Scilight

Transactions on Artificial Intelligence
https://www.sciltp.com/journals/tai

Article

CurateXelerator: A Collaborative Human-AI Framework for Accelerating Curatorial Practice

Qinyan Li and James She *

Computational Media and Arts Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511453, China
* Correspondence: jamespmshe@hkust-gz.edu.cn

How To Cite: Li, Q.; She, J. CurateXelerator: A Collaborative Human-AI Framework for Accelerating Curatorial Practice. *Transactions on Artificial Intelligence* **2025**, *1*(1), 291–306. https://doi.org/10.53941/tai.2025.100020

Received: 17 September 2025
Revised: 9 October 2025
Accepted: 23 October 2025
Published: 18 November 2025

Abstract: Curatorial texts are essential interpretive tools in art exhibitions, bridging the communication between artworks, curators, and visitors. While advancements in AI, particularly large language models (LLMs), have opened new possibilities for automating and assisting in the creation of curatorial texts, current AI models often suffer from inaccuracies and limited interpretive depth. This paper proposes "CurateXelerator", a collaborative Human-AI framework that integrates a structured input strategy into the curatorial workflow to address these challenges. Comprehensive evaluations demonstrate that proposed method generates texts significantly superior to baseline AI prompts in narrative coherence, rhetorical style, and adherence to constraints. Crucially, in a human paired-samples study, texts by CurateXelerator earned an average quality rating of 3.17/5, achieving statistical parity with texts from human writers, which averaged 2.89/5, while significantly outperforming them in technical simplicity and logical structure. This research contributes a validated collaborative Human-AI framework that enhances curatorial efficiency and quality, a reusable dataset and benchmark for future cutorial practice, and critical insights for integrating AI into Human-centered creative practices.

Keywords: curatorship; curatorial practice; AI; LLM; human-AI collaborations

1. Introduction

Curatorial texts play a critical role in art exhibitions, serving as essential interpretive tools that bridge communication between artworks, curators, and visitors. Effective curatorial materials not only provide visual analysis of artworks but also contextualize them within broader cultural, historical, and thematic frameworks. Major art institutions, such as the Smithsonian Institution and the Victoria and Albert Museum, advocate concise, engaging, and visitor-centric curatorial writing styles, emphasizing structured and standardized approaches to ensure clarity and coherence [1,2]. Recent advancements in artificial intelligence (AI), particularly large language models (LLMs), have opened new possibilities for automating and assisting in the creation of curatorial texts, potentially enhancing curatorial efficiency and interpretive quality. However, despite these promising developments, significant challenges remain unresolved. Current AI models frequently exhibit hallucination issues, generating inaccurate or exaggerated content when performing visual analysis or interpreting [3–5]. Moreover, practical curatorial texts require more than accurate visual descriptions; they must also address historical context, creative background, emotional resonance, and thematic relevance. Existing AI tools, while useful for preliminary research, remain inadequate for generating final, publishable curatorial texts due to limitations in interpretive depth, accuracy, and adaptability to nuanced curatorial judgments. Thus, a clear research gap exists between current AI capabilities and the practical needs of curatorial text creation. To bridge this gap, the following research questions are aimed to address in this work:

(a) How can a structured input strategy mitigate AI hallucination and enhance interpretive depth in curatorial text generation? This guides the design of our LLM-based generation method in Section 4, which proves effective in mitigating AI hallucination and enhancing interpretive depth.
(b) To what extent does the proposed framework achieve human-level quality in curatorial materials, as measured by multi-dimensional benchmarks? This motivates our validation strategy in Section 5, which validates the ability of framework to achieve human parity while exceeding expert performance in structural clarity.

Therefore, this work proposes a novel collaborative Human-AI framework for accelerating curatorial practice, CurateXelerator. This framework systematically integrates AI capabilities into the traditional curatorial workflow. It typically involves 3 main processes: preparation, generation, and evaluation. In addition, as shown in Figure 1, CurateXelerator also includes 4 key components:

(1) A trainable dataset designed to enable LLMs to effectively learn and adapt to varying exhibition language styles and maintain consistency across curatorial materials.
(2) A clearly defined input strategy that specifies the type and detail level of information, forming the foundation for effective AI-assisted text generation.
(3) An LLM-based generation method that strategically matches different detail levels of input information according to specific curatorial application scenarios.
(4) An evaluation benchmark developed to systematically validate the effectiveness of the proposed generation method and provide a reference standard for future research.

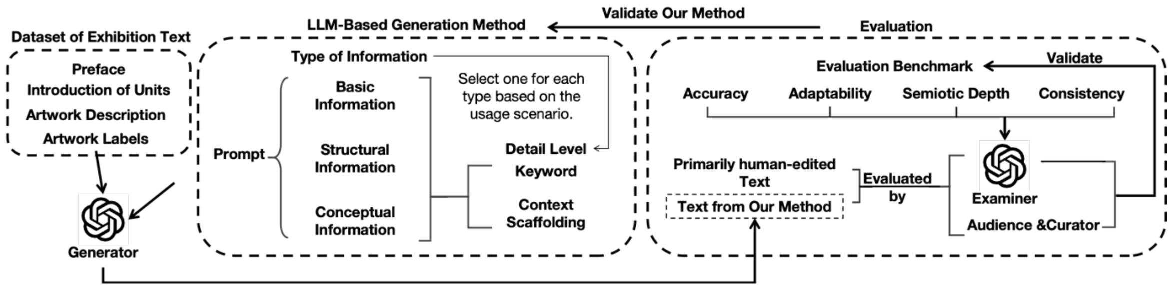

Figure 1. Overview of CurateXelerator: the Framework, 3 processes and 4 components.

The contributions of this research are threefold. First, it provides a practical and systematic AI-assisted curatorial framework, CurateXelerator, that significantly enhances curatorial efficiency, reduces the workload associated with frequent revisions, and improves the interpretive quality of curatorial texts. Second, it offers a structured, trainable dataset and a validated evaluation benchmark, both of which facilitate continuous improvement and future research in AI-assisted curatorial practices. Third, through comprehensive evaluations involving both LLM-based metrics and human feedback, this strategy demonstrates the rationality, effectiveness, and learnability of the proposed framework, laying a solid foundation for further exploration of AI role in the curatorial practice.

2. Related Work

This section reviews literature from two intersecting domains: human curatorial standards and the application of AI in creative writing. In the following content, the guidelines from authoritative curatorial institutions will be reviewed and organized. This will help to understand what characteristics a good curatorial document should have and what aspects to pay attention to when preparing a curatorial document. This part of the content will also guide the benchmarks proposed in the following section. Then current AI tools will be assessed against this standard, identifying a critical research gap: they often hallucinate and fail to address the holistic process of curatorial writing beyond simple visual analysis. This review motivates our framework, designed to bridge the gap between narrow AI capabilities and demands of curatorial practice.

2.1. Curatorial Material

Analyzing key considerations of human curators helps develop trainable frameworks for AI-generated texts. Although exceptional interpretive writing may transcend standardized structures, empirical studies indicate novices—including AI systems—benefit initially from standardized templates to accelerate learning and ensure baseline effectiveness [6]. For AI, this reduces risks of incoherence or contextual misalignment during early training, laying a foundation for later adaptability [7]. Thus, this work summarizes the common requirements of human approaches to the preparation of curatorial materials in this section.

There are two common principles of curatorial materials. The first common principle is the emphasis on visitor requirements. Although some critics argue curatorial practice should prioritize intentions from creators over visitor needs, this viewpoint is contested. Exhibitions inherently facilitate intellectual and cultural exchange, presupposing communication between creators and audiences. Thus, guidelines from art institutions [1,2,8–11] consistently identify visitor needs as fundamental design parameters. A second core principle is the adherence to a consistent writing style and structure. The Smithsonian Institution advocates a concrete, concise, relatable, and engaging style [1]. The Victoria and Albert Museum recommends a hierarchical structure: first layer—topics and key messages; second layer—themes and supporting stories; third layer—details and provenance [2]. Accordingly, interpretive texts prioritize essential information at the outset, aligning with audience needs and hierarchical principles.

2.2. AI for Creative Writing

Recently, certain scholars and artists have initiated explorations into the application of AI in curatorial practices through both artistic experimentation and academic research. Zheng et al. recently proposed ArtMentor [12], an AI-assisted evaluation tool for artworks, which can assist art educators in evaluating and providing guidance on student artistic works. This demonstrates the feasibility of using AI to analyze art. However, in the process of using AI to generate texts, research has shown that LLMs will exhibit hallucination [3,6], while Wang et al. [4] and Lu et al. [5] have shown that LLMs can also exhibit hallucination in the context of visual analysis. This poses a significant challenge to the use of AI in curatorial text creation, as exhibition texts are primarily used to provide written interpretations of the visual symbols within artworks. Chakrabarty et al. [13] identified common flaws in AI-generated creative writing, like clichés and purple prose, and fixed them using a taxonomy, human-edited dataset, and automatic tools to better align AI with human writing; however, they did not study hallucinations due to their focus on fiction, and editing can still introduce factual errors as noted in prior work [14]. Motivated by these challenges, a large multimodal model was proposed for painting analysis and creation, which was trained on a large dataset of paintings [15]. This breakthrough achievement demonstrates that the quality of texts generated by GalleryGPT is much better than baseline models. The training process significantly reduces LLM hallucination, thus ensuring the accuracy of the output text, allowing for a more precise description of the objects, colors, and structures within the artwork.

To address these challenges in collaborative AI-assisted writing, Dhillon et al. [16] explored varied scaffolding levels to enhance objective writing quality, such as story coherence, originality, and structure, though this may compromise subjective user satisfaction, including aesthetic appeal and sense of control. Similarly, Mirowski et al. [17] introduced the Dramatron system, which uses hierarchical generation and professional evaluations to enable LLM-assisted co-writing of coherent scripts, tackling long-form narrative semantic coherence, but hallucinations persist as a limitation. Furthermore, Anderson et al. [18] conducted user studies comparing ChatGPT with non-AI tools for creative ideation, finding that LLMs increase the quantity and detail of ideas but lead to homogenization at the group level. Complementing these efforts, Lee et al. [19] proposed a design space for intelligent writing assistants, structured around task, user, technology, interaction, and ecosystem aspects from a review of 115 papers, providing a unified framework to bridge fragmented research, though it requires ongoing updates for future advancements. Additionally, Laban et al. [14] developed the InkSync system, which offers executable edit suggestions and a Warn-Verify-Audit framework for transparent text editing, addressing lacks in author control and factual accuracy in standard chat interfaces, but with limitations like short study duration, small document scales, and reliance on specific LLMs.

However, in practical exhibition contexts, artwork descriptions are not limited to visual analysis. In addition, a comprehensive artwork description should include, but is not limited to, the background of creation, the story behind it, the emotions it seeks to convey, and its connection to the exhibition theme. Moreover, in addition to descriptions of the artworks themselves, exhibition texts may also need to encompass other curatorial materials, such as broader curatorial statements that go beyond artwork analysis, and more abstract and simplified labels for the works. In response to the aforementioned research gaps, this paper will continue find more details about current issue and challenge in curation area especially that exists when using AI to assist with curatorial material.

3. Challenges and Insights

To ground our theoretical understanding in real-world practice, Semi-structured interviews were conducted with professional curators. The goal was to identify the primary pain points in contemporary curatorial workflow and to understand the specific barriers to effective AI adoption. This section synthesizes our findings, first outlining the general challenges of time pressure and collaborative friction, followed by an analysis of the specific shortcomings curators encounter with current AI tools, such as conceptual exaggeration and lack of controllability.

https://doi.org/10.53941/tai.2025.100020

These real-world insights collectively reveal a critical need for a more structured and reliable system, directly motivating the design of the CurateXelerator framework.

Challenges in the Curatorial Workflow: Curators, whether working individually or collaboratively, encounter significant challenges in their curatorial workflow. Individual curators, such as Curator 3, face intense time pressure, managing numerous tasks and responsibilities on their own. For instance, Curator 3 had to curate two exhibitions in less than half a month, demonstrating the efficiency achievable with a streamlined process. In contrast, collaborative curators grapple with extensive team communication and coordination. Curator 1 noted that during a strongly curated, open-call exhibition, repeated adjustments to works of artists and the exhibition theme required extensive team communication, extending the preparation time to half a year. These challenges are further compounded by the interconnected and frequently revised nature of exhibition texts, which can trigger a cascading effect of further revisions.

Issues with AI: Curators encounter specific limitations when using AI in curatorial practice. While AI is seen as useful for research tasks such as gathering literature and identifying key concepts, it is not considered suitable for generating final curatorial texts. Curator 1 emphasized that AI can exaggerate concepts and struggle to define new ideas clearly, requiring careful management by human experts to maintain intellectual rigor. Curator 2 highlighted that role of AI should be goal-oriented, assisting in identifying relevant exhibition concepts and visual symbols, but not replacing human judgment in balancing commercial and academic priorities. These limitations of AI highlight the need for a more efficient approach to curatorial text creation.Finally, curation needs to focus on the connection between the exhibition content and the location of the exhibition to meet local characteristics, avoid cultural misunderstandings, and promote cross-cultural representation [20,21]. Studies have demonstrated that it is difficult for artificial intelligence to generate text at this stage, as LLMS often inevitably exhibit biases [22,23].

Motivations: These challenges and insights from interviews reveal a key pain point in curatorial work: the combination of frequent textual revisions and internal team coordination significantly increases workload, negatively affecting curatorial efficiency and quality. This has led to the motivation for proposing a new input strategy. Compared to traditional manual text editing, the proposed strategy aims to leverage AI to help the curatorial team work more efficiently while ensuring that the text is more usable than the normal prompt input method.

4. CurateXelerator: Proposed Collaborative Human-AI Framework for Accelerating Curatorial Practice

To generate high-quality curatorial materials, curators typically follow a structured workflow involving 3 main processes: preparation, generation, and evaluation. Therefore, our proposed framework, CurateXelerator, is designed around these 3 processes and consists of 4 key components: A trainable dataset, the types and detail levels of input information, the LLM-based generation method, and the evaluation benchmark. The first part, a trainable dataset is designed to enable LLMs understand the changes of different exhibition language styles and the consistent linkage of different curatorial materials. The second part distinguishes the information needed in the input process from the types and detail levels, which is the basis of the proposed method. And then the above elements will structured the proposed LLM-based generation method, advising to match different detail levels of input information based on application scenarios. Finally, an evaluation benchmark used not only to validate our method but also to provide a standard for future research.

4.1. Dataset

To ground our framework in real-world curatorial practice, a comprehensive dataset of 50 sets of professional curatorial materials were collected. The collection was intentionally diversified to represent a wide spectrum of exhibition objectives and narrative styles. It is crucial to clarify the role of this dataset: as detailed in Table 1, it served as a rich knowledge base for in-context learning, enabling the LLMs to understand and replicate the structures, styles, and terminologies of professional curatorial writing. This approach allows our framework to remain model-agnostic while still leveraging high-quality, domain-specific examples.

To ensure the quality and relevance of this collection, the dataset was constructed through a systematic, multi-stage filtering process, as illustrated in Figure 2. This procedure began with a broad identification of 103 potential exhibitions, which were subsequently screened based on data accessibility and the qualitative rigor of their curatorial texts. The final inclusion of 50 exhibitions was determined by a thorough quality review, ensuring that all texts met key criteria for readability, completeness, and logical coherence.

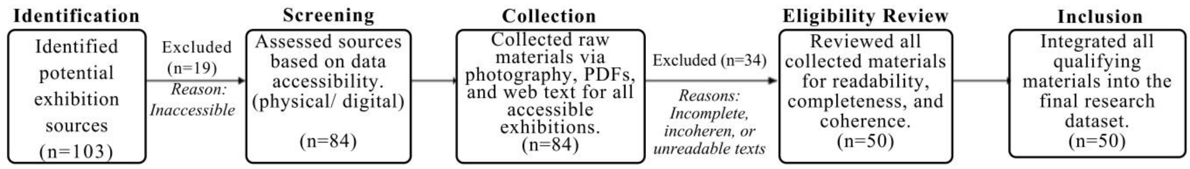

Figure 2. Exhibition filtering process for Dataset curation.

Table 1. Specifications of the curatorial dataset.

Specification	Description
Text Components	Curatorial Statements (Forewords/Afterwords), Unit Introductions, Artwork Descriptions, and Artwork Labels.
Languages	English and Chinese.
Primary Purpose	To serve as a knowledge source for in-context learning, providing the LLMs with structural and stylistic exemplars for text generation.

The rationale behind our collection strategy was to capture how curatorial texts adapt to different exhibition goals. The structure and tone of a text for a sales-oriented exhibition, for example, differ significantly from one written for an academic, theme-driven showcase. Therefore, the collected sets were categorized based on their primary curatorial objectives. This goal-oriented classification, summarized in Table 2, ensures that the LLMs can learn to generate contextually appropriate texts by drawing on relevant examples during the prompting process. The thematic-driven exhibitions form the largest category, reflecting the prevalence of conceptually grounded curatorial practices in contemporary art. The complete dataset, which represents the logical relationships between exhibition themes and individual artworks at both macro and micro levels, is available for review and future research at the following link: https://drive.google.com/drive/folders/1lwCaoK8hcOPXusLyx7nfPzqv2nGah1cf?usp=sharing.

Table 2. Distribution of exhibition types in the dataset.

Type of Exhibition (by Objective)	Number of Sets
Thematic Driven Exhibitions	21
Commemorative Exhibitions	15
Educational & Public Outreach Exhibitions	6
Sales-oriented Exhibitions	4
Community & Social Engagement Exhibitions	4
Total	50

4.2. Type and Detail Level of Input Information

This section defines our input strategy, which classifies information along two dimensions: information type and detail level. These dimensions form the input template for our generation method. This input strategy is inspired by the The "Outline, Subject, Form" curatorial trilogy to find topic of curation proposed by [24]. The template is proposed to include three types of information: basic information, structural Information and Conceptual Information. Table 3 provides detailed description of these concepts.

Basic Information captures individual information for each material type, including the name of exhibition, unit, artwork, artwork pictures, creators, visual elements, and text style. Structural Information reflects inter-material relationships from the whole to the individual. Conceptual Information embodies the collective tone and overarching commonality of all materials, integrating theoretical, philosophical, cultural concepts, vision of exhibition, and type of exhibition. This layer operates as a cultural coding system.

Given the interdependence of the three types, Detail Level is proposed to quantify input information across methodologies. Rather than assuming more input necessarily yields better output, this study aims to identify an efficient input method that produces high-quality texts with minimal preparation for specific scenarios. As information density affects preparation time, two detail levels are defined: Keyword and Contextual Scaffolding. Keyword highlights coreterminologies, while Contextual Scaffolding integrates comprehensive contextual details.

Table 3. A table showing types and detail levels of input information with examples.

Type of Input Information	What it Contain	Detail Level	Examples in Tanka AI Art Exhibition
Basic Information	Name of exhibition, unit, and artwork, Type of artwork, Belonging unit, Picture of artwork, creator(s), Visual elements	Keyword	"EVEBOUND": Painting, Ink Series, Fisherman, TankaPeople, Tanka Boat, Tanka Hut (Water House)
		Context Scaffolding	"EVEBOUND", Ink Series Painting: Tanka people silhouette at twilight, AI-enhanced light and shadow, symbolizing tradition and modernity.
Structural Information	Structural requirement, Reasons for dividing units, Style of text	Keyword	Foreword: Three-part structure, Introduction, Artistic Method, Conclusion.
		Context Scaffolding	Foreword divided into three parts: Tanka culture back-ground, AI technology use, exhibition significance.
Conceptual Information	Type of exhibition, Vision of exhibition, Related theoretical, philosophical or cultural concepts	Keyword	Cultural Concepts: Tanka Culture, AI Art
		Context Scaffolding	Tanka people: traditional lifestyle fading, AI and new media revitalize cultural presentation.

4.3. LLM-based Generation Method

Our proposed method involves preparing the three types of input information at two distinct detail levels in collaboration with AI according to different scenarios of text usage. Figure 3 shows an example of how a curator might work on a text using the proposed method. For instance, with the aim of writing the exhibition text for a public art project, it would be better to input basic information with the "Context Scaffolding" description, structural information and conceptual information with "Keyword".

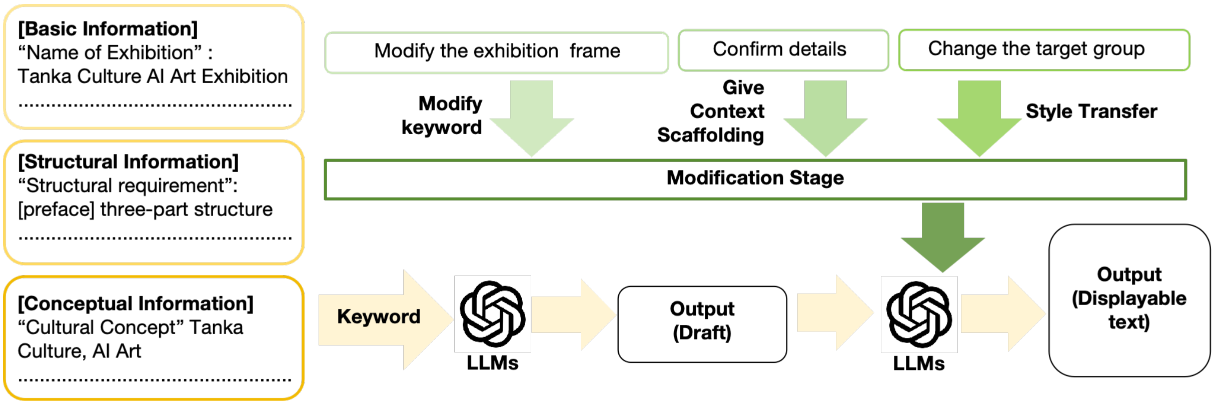

Figure 3. Schematic diagram of LLM-based Generation Method.

The following section (Section 5) will conduct a an experiment to compare our proposed methods against a baseline. Two methods will be proposed suitable for different scenarios using the proposed input strategy. They are Method 1 (M1) for draft and Method 2 (M2) for displayable text in the test, while Method 0 (M0) is the baseline without our method. These three methods were tested in the curatorial preparation, including curatorial statements, introduction of unit, artwork description, and artwork label for exhibition held by our team and another 2 exhibitions from other institutions. To validate the LLM-based generation method, the generated texts are then evaluated by audiences, experts, and LLMs using the benchmark detailed in Section 4.4, with the full results presented in Section 5.

4.4. Evaluation Benchmark

To systematically evaluate the quality of AI-generated curatorial materials, a multi-dimensional benchmark framework will be proposed based on the guideline from authority institution and interviews from curators. This framework is composed of four key dimensions, whose proportional weights are visualized and whose specific evaluation criteria are detailed in Figure 4.

The cornerstone of this benchmark is Content Accuracy, which is assigned a dominant weight of 50% to directly address the critical issue of AI hallucination and uphold the academic credibility essential for museum contexts. This dimension is bifurcated into two tiers: core facts, which demand a zero-error standard for foundational information like artwork titles and dates to maintain audience trust, and professional expression, which ensures academic rigor through the precise use of terminology. This hierarchical design reflects the museological consensus on balancing factual correctness with intellectual depth [11,25–27].

The remaining 50% of the evaluation is distributed across three dimensions that assess communicative effectiveness, structural coherence, and interpretive depth. Scenario Adaptability evaluates a practical effectiveness of text by assessing its readability for non-specialist audiences and its suitability for the specific exhibition context. To ensure narrative integrity across the entire exhibition, Logical Consistency is assessed through both horizontal consistency (ensuring uniform terminology for related items) and vertical consistency, maintaining thematic alignment from the overall curatorial statement down to individual labels. Finally, Semiotic Depth measures the higher-order interpretive power of text, evaluating its capacity for symbolic decoding to prevent cross-cultural misinterpretation [20,21] and its ability to foster empathic resonance by creating meaningful cognitive and emotional connections between the artwork and the visitor.

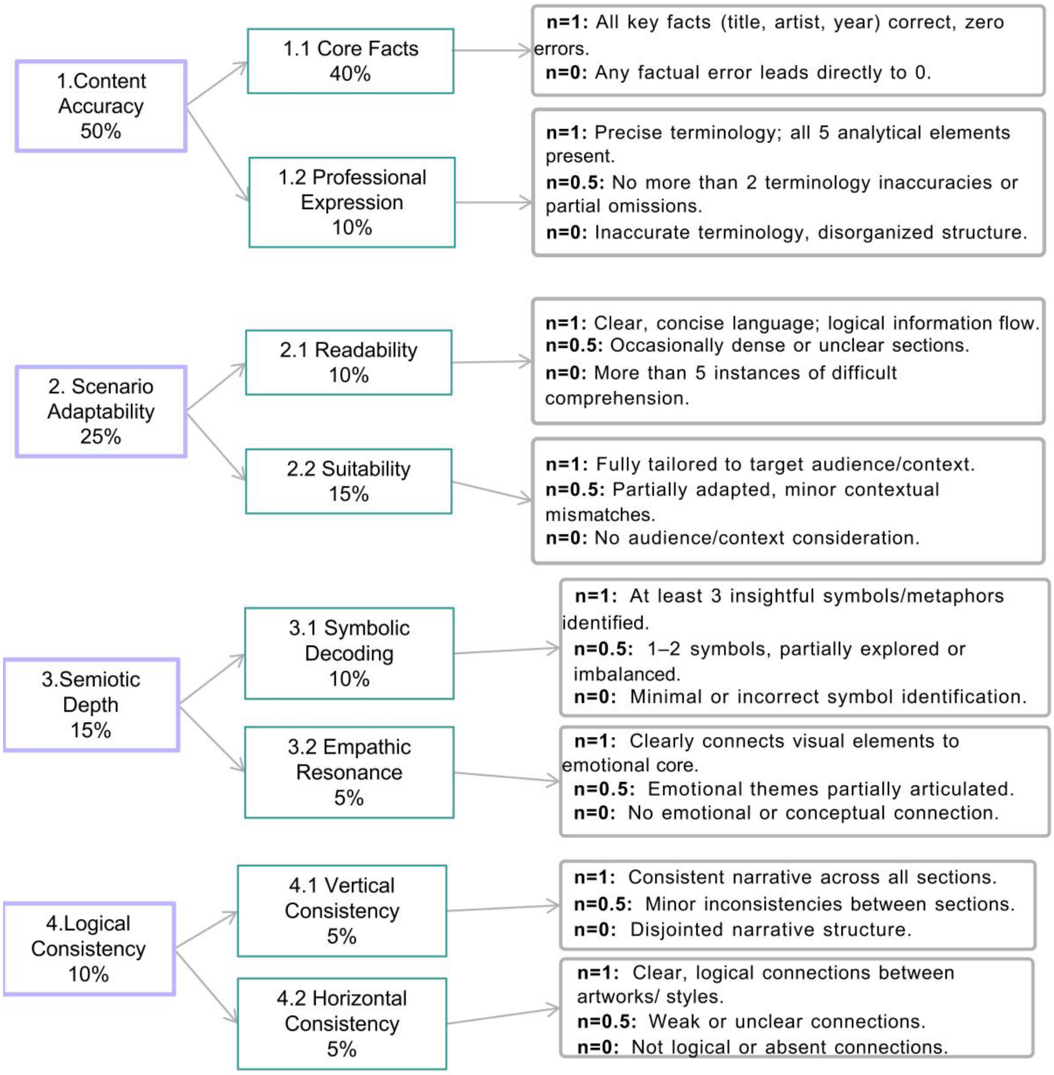

Figure 4. Proportional weights and visual representation of the evaluation benchmark.

5. Validation

This section details the multi-pronged validation strategy designed to rigorously assess the CurateXelerator framework. Our evaluation follows a deliberate progression from qualitative observation to quantitative proof. We begin with a qualitative analysis to first establish the nature of the improvements our structured method (M2) offers over a standard baseline (M0), highlighting fundamental differences in narrative structure, style, and depth. Building on these insights, a two-phase quantitative validation will be conducted. The first phase employs a scalable LLM-based evaluation to objectively measure the performance gradient across all methods (M0, M1, M2) and benchmark them against human writing. The second and definitive phase consists of a formal human evaluation, where a paired-samples test is used to statistically compare our framework output directly against texts from human experts. This multi-phrase validation strategy allows us to not only demonstrate that our method is effective but also to precisely articulate how it excels and to validate its ability to achieve human parity. In the test, the LLMs used in the generation process are GPT-4o [28] and LLMs used in the evaluation process are GPT-4o [29], Gemini 2.5 Pro [30], DeepSeek [31].

5.1. Validation Design: Baselines and Proposed Methods

The three experimental methods—a baseline (M0) and two framework-based approaches (M1, M2)—are compared in Table 4. The baseline, M0, simulates standard unstructured prompting by consolidating all source text into a single paragraph, establishing a benchmark without the framework's structured guidance. In contrast, M1 operationalizes the framework for rapid ideation, utilizing concise inputs like keywords and bullet points. M2 represents the framework's full potential, employing detailed descriptive sentences and structural constraints to generate publication-ready texts. This three-method design ensures a systematic and fair comparison. By holding the LLM and source information constant, any variance in output quality can be attributed solely to the input's structure and richness. The progression from M0 to M2 thus allows for isolating the effects of introducing structure (M1 vs. M0) and subsequently enriching it with detailed context (M2 vs. M1).

To validate real-world effectiveness, these methods were applied to two distinct exhibition scenarios: one focused on Tanka Culture AI Art [31] and another on sustainable soil landscapes, as detailed in Tables 5 and 6. These tables illustrate the structural shift from unstructured input of M0 to template of M1. The M2 prompt, not shown separately, shares structure of M1 but enhances it by replacing keywords with fully elaborated sentences.

Table 4. Comparison of Generation Methods (M0, M1, M2) in the Validation.

Aspect	M0	M1	M2
Core Principle	Simulates a "naive user" approach by providing all exhibition information as a single unstructured block of text.	Implements the proposed framework using concise keywords and essential bullet points.	Fully implements the proposed framework with detailed, sentence-level contextual information across all template tiers.
Structure	Unstructured narrative paragraph.	Utilizes the 3-tier template (Basic, Structural, Conceptual).	Utilizes the same 3-tier template populated with rich contextual sentences.
Conceptual Input	"...The exhibition discusses Tanka culture's disappearance and uses AI to preserve it..."	Cultural Concepts: Tanka culture, AI Art Vision: Preserve Tanka culture	Cultural Concepts: "Tanka people are boat-dwelling residents..." Vision: Use AI to re-present ancient cultures in a global context
Structural Input	None explicitly provided.	Structural Requirement: Three-paragraph foreword (700–800 words).	Structural Requirement: Foreword with three sections—Intro (150–200 words), Methods (300–400 words), Conclusion (100–200 words).
Intended Use Case	Establishes a performance benchmark without structured guidance.	Early-stage curation: quickly generate drafts for discussion without heavy input.	Final-stage curation: produce publication-ready texts with maximum fidelity to curatorial intent.

Table 5. Prompt of different methods in Exhibition 1.

Method	Example
M0 Unstructured (Baseline)	Please write curatorial text for the "Greater Bay Area Cultural Series—Tanka Culture AI Art Exhibition." The exhibition includes the Ink Series, Meticulous Series, and Comic Series, featuring works such as "EVEBOUND", "AQUA VOWS", and "FADING HOME". These pieces are created by..., and "EVEBOUND" shows the daily fishing routine of this ethnic group...(details of each piece).... This exhibition aims to promote and preserve Tanka culture through AI art. The text needed includes: an exhibition foreword, introductions for each unit, descriptions of each artwork (in the character of an old Tanka man),and artwork labels. The foreword, unit introductions, and labels should be formal and academic, while the artwork descriptions should be vivid and conversational.
M1 Structured (Keyword Level)	Task: Generate a preface of the exhibition named "Greater Bay Area Series—Tanka Culture Artificial Intelligence Art Exhibition", introduction of each unit, and art label. (1) Basic Information: • Exhibition name: "Greater Bay Area Culture Series—Tanka Culture Artificial Intelligence Art Exhibition" • Units: Ink Series, Meticulous Series, Comic Series

Table 5. *Cont.*

Method	Example
M1 Structured (Keyword Level)	• Artworks: EVEBOUND (ink painting: fishermen, Tanka people, boats, sheds); AQUA VOWS (comic: Tanka boat, red silk); FADING HOME (comic: Tanka shed, boat, ocean) • Creators:...... • Types: foreword, unit intros, artwork descriptions, labels • Styles: foreword/units/labels formal/academic; descriptions formal spoken (tour guide grandpa perspective) (2) Structural Framework: • Foreword: Three-part structure: introduction and premise, artistic method and analysis, macro significance and conclusion (700–800 words) • Unit intros: Unit group name, type/style, relationship to exhibition, rule/project association, commonality analysis (300 words) • Artwork desc.: Work name, creator and time, core concept/theme association, visual elements/structure/color, creation details (300–350 words) • Label: Title, author, time, group identifier, background introduction, detail description (50 words) • Unit divisions: Ink (early history, colorless, labor scenes); Meticulous (stable life, fine colors, family themes); Comic (new media transformation, ashore life, desolate atmosphere) • Artworks per unit: Ink (Rowing, EVEBOUND); Meticulous (FUN TIME, DREAM-ING); Comic (AQUA VOWS, FADING HOME) • Stories: EVEBOUND (homecoming snapshots, work and reunion); AQUA VOWS (red boat, empty scene, traditional wedding, empty house); FADING HOME (stilt houses, deserted scenes, traditions gone) (3) Cultural Semiotics (Keywords): • Concepts: Tanka Culture, AI Art • Vision: Disseminate and protect Tanka Culture • Type: Cultural promotion and academic exhibition Generate text adhering to structure.

Table 6. Prompt of different methods in Exhibition 2.

Method	Example
M0 Unstructured (Baseline)	Tone: Fluid, lyrical, reflective. Exhibition Title: Flowing Circulation—Co-creating a Cyclical Soil Landscape. Brand: Zhao Ri Wei Pin. Concept: Soil as vitality symbol and co-creation medium; endless renewal cycle. Content: Interactive “soil slices” from farm feed; visitors participate in landscape co-creation. Style: Poetic, organic rhythm; natural imagery. Length: 400–600 words. Generate text based on above.
M1 Structured (Keyword Level)	Task: Write exhibition text for “Flowing Cycle—Co-creating the Landscape of Circular Soil” (Vipshop Milk brand), including preface and one unit introduction. (1) Basic Information: • Title: Flowing Cycle—Co-creating the Landscape of Circular Soil • Unit: Flowing Cycle Slice Pasture • Artwork: Soil Slice from dairy feed • Creator:... • Visuals: Natural tones, rustic style • Type: Interactive installation • Text Style: Poetic, philosophical (2) Structural Framework: • Preface: Soil philosophy, cyclical flow, brand practices, participation invite • Unit Framework: Theme (Pasture); Artwork origin/story; Brand meaning; Interaction (Take slices to co-create) • Division Reasons: Brand chain (soil to milk); Cyclical visualization (3) Cultural Semiotics (Keywords): • Concepts: Cyclical flow, human-nature symbiosis, sustainability, mindful lifestyle • Vision: Visualize philosophy, promote interaction, advocate eco-living, invite participation Generate text adhering to structure.

5.2. Qualitative Analysis: Baseline and Proposed Methods

To illustrate the practical impact of our structured prompting framework, this section will begin with the qualitative analysis. Our objective is to identify and articulate the specific linguistic, structural, and stylistic differences between texts generated using a standard baseline prompt (M0) and those generated via our proposed rich-context method (M2). To ensure the robustness of our findings, two distinct exhibition cases were focused for this analysis: the "Tanka Culture AI Art Exhibition", shown in Figure 5, which focuses on cultural heritage, and the "The flowing cycle—Jointly building a circular soil landscape", shown in Table 7 a speculative case centered on technology and society. While their themes differ, both cases share the core curatorial goal of engaging and informing the audience. Our qualitative methodology was centered on in-depth, semi-structured interviews conducted as a secondary evaluation phase. Professional curators and the general public with an interest in arts and culture were invited to evaluate the versions from M0, M1 and M2 for 2 exhibitions showed in Figure 5 and Table 7.

	Preface - First Paragraph - Motivation Illustration
Ours (M2)	The waters of the Guangdong-Hong Kong-Macao Greater Bay Area have nurtured the unique Tanka culture. This "Gypsies of the Sea" community lives on boats and coexists with the tides, embodying philosophy of...... In the process of urbanization and modernization, Tanka stilt houses have gradually been replaced by concrete structures, and waterborne wedding customs have given way to land-based ceremonies. The cultural memory of this millennia-old community is facing irreversible disappearance...... By using.....,we construct a cultural memory field that blends the virtual and the real.
Baseline (M0)	The "Greater Bay Area Series - Tanka Culture AI Art Exhibition" transforms this vanishing waterborne memory into an eternal spiritual totem through art innovation empowered by digital technology. The exhibition, centered around an interdisciplinary team, completed the technical development and artistic creation using AI technologies such as generative adversarial networks and style transfer algorithms, ...within 3 years.
Human Writing	The Tanka people are water-dwelling inhabitants of China's southeastern coastal regions. Born on rivers, they ...... However, with the process of modernization, the Tanka lifestyle and traditional customs are gradually being forgotten. Through new technologies for media and art, ... By leveraging the text processing......, we have been able to uncover the historical context of the Tanka people, recreate historical scenes....... This allows audiences in the Greater Bay Area, as well as nationwide and globally, to see Tanka culture, promoting cultural identity and protecting cultural diversity.

	Introduction of Units - Cartoon Series	Artwork Label		
Ours (M2)	**Cartoon Series** The cartoon series represents the impact of modernization and digitalization, showcasing the life of the Tanka people after "coming ashore" through digital tools. No longer drifting fishermen, the Tanka people are now settled on land, depicting the context of this historical transition. Through empty scenes devoid of characters and cold-colored imagery, the series reflects the disappearance and loss of traditional culture with deserted fishing boats and desolate waters.	**EVEBOUND** xxx Team 2024 **Ink Painting Series** Dynamic ink recreates the Tanka stilt house communities, with AI algorithms capturing the changing tides and shadows, while the ink's resonance narrates the epic of a thousand years of boat dwelling.	**AQUA VOWS** xxx Team 2024 **Cartoon Series** Surreal red boats float in cold-colored waters, with scattered pixels witnessing the symbolic dissolution of Tanka wedding customs in the wave of modernization.	**FADING HOME** xxx Team 2024 **Cartoon Series** A dialogue between mortise and tenon joints and digital noise, with blue-green landscapes expressing the pain of identity reconstruction as Tanka people transition from water to land.
Baseline (M0)	**Cartoon Series "Waterfront Chronicles"** Utilizing natural language processing technology, oral histories are transformed into visual narrative scripts. AI automatically matches storyboard styles based on emotion analysis results, preserving the essence of traditional comic composition while incorporating the temporal leaps of cinematic montage.	**EVEBOUND** xxx and AI System 2024 Generative Adversarial Networks, Mineral Pigments, Raw Rice Paper 45 × 180cm Dynamic Ink Installation	**AQUA VOWS** xxx Multimedia Laboratory 2024 Parametric Design, Art Micro-Spray, Silk 70 × 70cm Augmented Reality Version	**FADING HOME** Digital Humanities Team 2024 3D Scanning, Algorithmic Generation, Traditional Mounting 200 × 80cm Interactive Image Installation
Evaluation	Green: Performs well, meets curatorial material requirements. Yellow: Contains some shortcomings, mainly due to excessive implementation details. Red: Wrong facts.			

Figure 5. Comparison of generated results with and without LLM-based Generation Method in Exhibition 1.

Table 7. Comparison of generated results with and without LLM-based Generation Method in Exhibition 2.

	Afterwords—Recall the Big Idea
Baseline (M0)	In nature's laws, cycles reign as primal wisdom. Zhaori Weipin initiates its 14-year soil rejuvenation initiative, tracing milk's journey to map symbiosis between earth, pastures, and life. This immersive journey, themed "Cycle", guides you through microbial-rooted subterranean forests, decodes grass-growth energy matrices, and witnesses cows' gentle reciprocity with the land they nourish.
Ours (M2)	Soil is life's cradle, time's layered archive, and a hidden verse of human-nature symbiosis. It embraces seeds, decomposes foliage, incubates life, silently weaving a web of coexistence. Through cultivation, sowing, and harvest, humans tap earth's rhythm, while soil responds with cyclical cadence—a millennia-spanning dialogue of interdependence, reciprocity, and perpetual flow.
Human Writing	Our brand views soil as life's cradle—metaphorical, not just earthly, but an ecosystem woven from sunlight, breezes, animal vitality, and human cultivation. This co-woven "soil" serves as both creator and creation. It perpetually nurtures new life through transformation, which in turn reshapes its surroundings, sparking fresh cycles of growth. This is nature's eternal flow.

They were asked to perform a comparative evaluation, articulating the perceived differences in quality, style, structure, and overall effectiveness. From their detailed feedback, the analysis was synthesized across four key dimensions: Narrative Structure, Rhetorical Style and Lexicon, Information Depth and Synthesis, and Controllability and Adherence to Constraints. To provide concrete textual evidence for our findings, examples from the "Tanka Culture AI Art Exhibition" (Figure 5) will be primarily shown. The core findings, which were consistent across both cases, are summarized in Table 8. This side-by-side comparison reveals a stark contrast between the outputs, highlighting the transformative effect of our framework.

Table 8. Qualitative Comparison of Texts from Baseline (M0) and Proposed (M2) Methods.

Analysis Dimension	**M0: Baseline (Unstructured Prompt)**	**M2: Proposed Method (Structured Prompt)**	**Key Insights & Connection to User Feedback**
Narrative Structure	Disjointed and linear; often reads like a list of facts. Fails to establish a clear narrative arc or purpose in the opening.	Coherent and thematically driven; opens with a strong hook and follows the specified curatorial structure with a clear beginning, middle, and end.	Aligns with feedback from a curator (User 2), who noted that M2 "effectively emphasizes the motivation of exhibition in the opening paragraph". The framework directly enforces a strong, purposeful narrative.
Rhetorical Style & Lexicon	Generic, repetitive, and fact-based language (e.g., "The Tanka people are a group that lives on boats..."). Lacks an engaging or professional tone.	Evocative and metaphorical language (e.g., "...the 'Sea Gypsies' of Southern China, a community whose entire existence unfolds upon the water".). Adopts a more literary and professional tone.	Explains User 1's preference for M2's "more elegant language" and User 4's fascination with the "expressive and impactful phrasing" of "Sea Gypsies", which they felt was beyond the capability of typical AI.
Information Depth & Synthesis	Repeats surface-level information from the input materials without deeper connection or interpretation. The text lacks a central, unifying idea.	Synthesizes information from different parts of the input (e.g., cultural history, artistic themes, specific artworks) to create novel insights and a cohesive narrative.	The framework compels the LLM to process and integrate information from multiple sources, moving beyond simple summarization to genuine synthesis, a key requirement for high-quality curatorial writing.
Controllability & Adherence to Constraints	Frequently ignores or misunderstands implicit and explicit constraints like word count, tone-of-voice, and section-specific goals.	Precisely adheres to specified constraints, delivering text that fits the required length, tone, and purpose for each section of the curatorial narrative.	Demonstrates the framework's primary utility in providing fine-grained control over the generation process. This reliability is a critical need for professional applications where outputs must meet specific editorial standards.

5.3. LLM-based Evaluation by the Benchmarks

In this section, a total of 144 evaluation experiments using different LLMs were conducted to assess 12 sets of curatorial materials in four evaluation dimensions. The experimental results are presented in Figures 6 and 7.

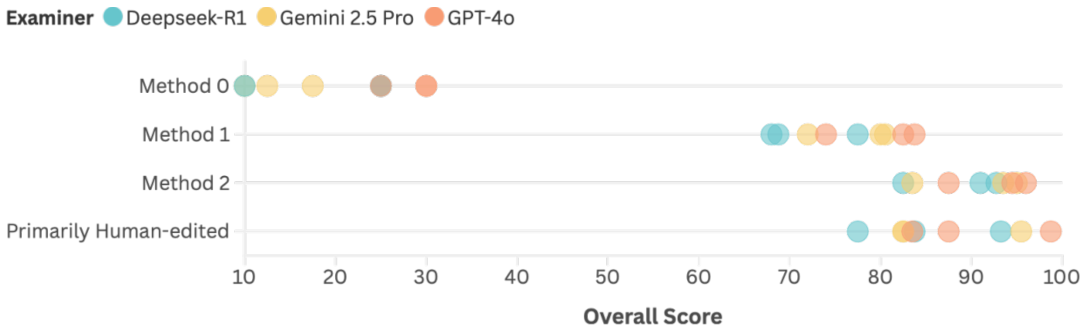

Figure 6. Comparison among the LLMs average scores in four dimensions obtained from three sets of curatorial materials using different methods to create (total score:100).

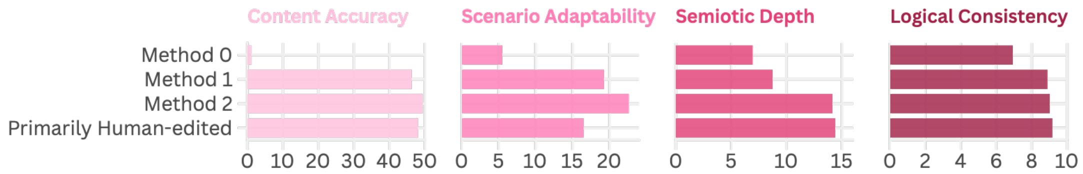

Figure 7. Comparison among the LLMs average scores in four dimensions obtained from three sets of curatorial materials using different methods to create (total score:100).

5.3.1. Goal 1: Validate Effectiveness of LLM-Based Generation Method

The evaluation results clearly demonstrate that methods employing our proposed structured input strategy (M1 and M2) significantly outperform the baseline method (M0), achieving scores close to human-written texts. Specifically, our strategy effectively improves Content Accuracy by mitigating hallucination issues common in LLM-generated texts, enhances Scenario Adaptability by enabling AI to adapt texts stylistically for different exhibition contexts, and consistently achieves strong performance in Semiotic Depth and Logical Consistency. Furthermore, the superior performance of M2 compared to M1, despite similar input volumes, confirms that structured input rather than merely increasing input quantity is crucial for improving text quality. Excessive input information may even negatively impact text quality and curatorial efficiency, highlighting the importance of carefully controlling input volume according to specific scenario requirements.

5.3.2. Goal 2: Validate the Learnability of Proposed Evaluation Benchmark

The scores assigned by different models to the same creation method exhibit a clear clustering trend, indirectly indicating that our proposed evaluation benchmark possesses strong learnability for LLMs. Meanwhile, the comparison between the evaluation criteria using LLM and the human scoring criteria in the following section can also verify this objective.

5.4. Human Evaluation with Questionnaires

While LLM evaluation provides a useful proxy, the definitive measure of quality is human perception. Therefore, a human evaluation study was conducted to provide the gold-standard validation for our framework. A total of 144 participants with diverse backgrounds were invited, and the general trends of this large-scale survey showed that texts generated by our M2 method consistently received high ratings (Figure 8). Meanwhile, the order of scores and the performance in each dimension of the human scoring results are consistent with those of the LLM scoring, which also proves the effectiveness of proposed evaluation benchmark.

For a more rigorous statistical analysis, a representative subset of 22 participants who identified as having a professional or academic background in the arts were isolated, as their domain expertise provides a more critical benchmark for evaluating curatorial writing. A within-subjects, paired-samples experimental design where each expert evaluated both a human-written text and a text generated by our CurateXelerator framework (M2) and the results can be checked in Table 9 were employed. To formally test our claims, the following null hypothesis (H1) were established:

H1: *There is no statistically significant difference in the perceived quality between human-written texts and those generated by our M2 method.*

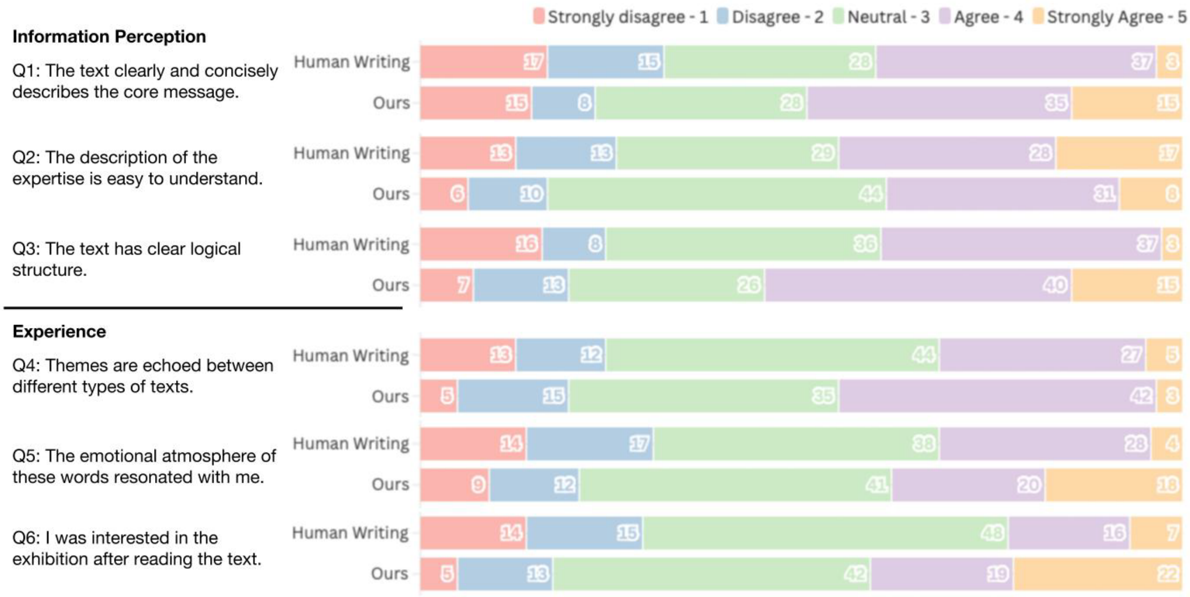

Figure 8. The distributions of scores in the questionnaire (total score:5).

Table 9. Paired Samplest-test Results for Human vs. AI-Generated Text Evaluations ($N = 22$).

Evaluation Dimension	Human Writing (M ± SD)	CurateXelerator (M ± SD)	*p*-value
Overall Score	2.89 ± 0.82	3.17 ± 0.69	/
Q1: Clarity & Conciseness	2.90 ± 1.41	3.29 ± 1.23	0.2137
Q2: Technical Simplicity	2.62 ± 1.16	3.33 ± 0.80	0.0097
Q3: Logical Structure	2.90 ± 1.09	3.52 ± 1.03	0.0239
Q4: Inspires Curiosity	3.05 ± 1.02	2.95 ± 0.80	0.7908
Q5: Evokes Emotion	3.05 ± 0.86	2.81 ± 0.51	0.1038
Q6: Sparks Interest to Visit	2.81 ± 1.12	3.10 ± 1.18	0.4468

5.4.1. Validation of Human Parity on Overall Quality

A dimensional analysis was conducted to assess the performance of the CurateXelerator framework against human-written texts. The results indicate that in the majority of evaluated dimensions, the AI-generated texts achieved a level of quality that was statistically indistinguishable from that of human writers.

Specifically, no significant differences were observed in four out of the six key dimensions. These include Clarity & Conciseness (Q21) ($p = 0.2137$), Inspires Curiosity (Q4) ($p = 0.7908$), Evokes Emotion (Q5) ($p = 0.1038$), and Sparks Interest to Visit (Q6) ($p = 0.4468$). The high p-values across these domains, particularly those related to creative and affective user experience (Q4, Q5), support the claim that the framework can produce content on par with human experts. While the AI-generated texts tended to receive slightly higher mean scores ($M = 3.17$) than human-written text ($M = 2.89$), this pattern demonstrates robust, human-level performance rather than a statistically significant overall superiority.

5.4.2. Beyond Parity: Identifying Areas of AI Superiority

While parity was established in most areas, a more granular analysis reveals key domains where the CurateXelerator framework demonstrated a significant advantage. The AI-generated texts ($M = 3.33$) were rated as significantly superior in Technical Simplicity (Q2), making complex information easier to understand, compared to the human-written texts ($M = 2.62$, $p = 0.0097$). Furthermore, they were perceived as having a more coherent Logical Structure (Q3) ($M = 3.52$) than the human-written texts ($M = 2.90$, $p = 0.0239$).

These findings suggest that the primary strength of framework lies in its ability to structure complex information into clear, digestible, and logically organized narratives. This advantage may stem from the structured input strategy, which compels the model to be more direct and systematic, effectively mitigating the "curse of knowledge" often seen in expert writing. This nuanced picture—achieving parity in creative and affective dimensions while demonstrating superiority in structural clarity—underscores the specific value proposition of integrating the CurateXelerator framework into curatorial practice.

5.5. Limitation

While the findings provide strong support for the CurateXelerator framework, we acknowledge several limitations that offer avenues for future research. The first limitation lies in the scope of our expert human evaluation; while the 22 arts professionals provided rigorous feedback, their demographic may not fully represent the diversity of global curatorial practices or the perspectives of non-expert audiences. Furthermore, the framework is also constrained to text-based inputs; extending its capabilities to accept multimodal inputs like images and diagrams remains a key area for future development. Addressing these challenges presents clear and promising directions for enhancing the generalizability and practical utility of framework in future iterations.

6. Conclusions

This paper introduced CurateXelerator, a human-AI collaborative framework designed to address the challenges of quality and controllability in AI-assisted curatorial writing. The research was guided by two primary questions, and the findings provide clear answers. In response to RQ (a), which asked how a structured input strategy could mitigate AI limitations, the results confirmed that the multi-tiered template of CurateXelerator significantly reduces factual errors and enhances interpretive depth compared to unstructured baselines. Regarding RQ (b), which questioned if AI could achieve human-level quality, the human evaluation provided definitive validation: the output ($M = 3.17$) of framework achieved a quality level statistically indistinguishable from that of human experts ($M = 2.89$) across the majority of creative and affective dimensions. Moreover, it significantly surpassed them in key areas of technical simplicity ($p = 0.0097$) and logical structure ($p = 0.0239$), demonstrating its capacity to not only match but also enhance professional curatorial writing.

The contributions of this work are threefold. First, this work provides a practical framework that demonstrably improves curatorial efficiency and text quality. Second, we contribute a reusable, categorized dataset and a multi-dimensional evaluation benchmark to advance future research in this domain. Third, the findings offer critical insights for designing effective human-AI collaborative systems in creative fields. Further research could also explore extensions of the framework, such as direct multimodal output and longitudinal deployment in real-world museum settings. Ultimately, CurateXelerator moves beyond simple automation to take a significant step toward realizing Human-Machine Symbiosis [32]. It establishes a new paradigm where AI tools function not as autonomous replacements, but as powerful, collaborative partners, thereby advancing human-centered creative practices.

Author Contributions

Q.L.: data curation, writing—original draft preparation, reviewing and editing; J.S.: supervision, reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability Statement

The related data presented in this study are openly available at: https://drive.google.com/drive/folders/1lwCaoK8hcOPXusLyx7nfPzqv2nGah1cf?usp=sharing.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used generative AI (LLM) to validate the proposed method and polish writing. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Smithsonian Institution. Guide to Interpretive Writing for Exhibitions. Available online: https://exhibits.si.edu/wp-content/uploads/2021/09/SI-Guide-to-Interpretive-Writing-for-Exhibitions.pdf (accessed on 18 September 2021).

2. Victoria and Albert Museum. Writing Gallery Text at the VA—A Ten Point Guide. Available online: https://www.vam.ac.uk/blog/wp-content/uploads/VA_Gallery-Text-Writing-Guidelines_online_Web.pdf?srsltid=AfmBOootqlWDBHpmM-NgNUrmUV11pcarcMSQAU_ANeWKRk8Aer-7jBuz (accessed on 1 November 2019).
3. Huang, L.; Yu, W.; Ma, W.; et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Trans. Inf. Syst.* **2025**, *43*, 1–55.
4. Banerjee, S.; Agarwal, A.; Singla, S. LLMs will Always Hallucinate, and We Need to Live with This. In *Intelligent Systems Conference*; Springer Nature: Cham, Switzerland, 2025; pp. 624–648.
5. Sriramanan, G.; Bharti, S.; Sadasivan, V. S; et al. LLM-Check: Investigating detection of hallucinations in large language models. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 34188–34216.
6. Brown, T.; Mann, B.; Ryder, N.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
7. Dathathri, S.; Madotto, A.; Lan, J.; et al. Plug and Play Language Models: A simple Approach to Controlled Text Generation. In Proceedings of the International Conference on Learning Representations, 26 April 2020–1 May 2020; pp.1–34.
8. The British Columbia Museum Association. The Art of Writing for Exhibitions. Available online: https://museum.bc.ca/wp-content/uploads/2020/12/BCMA-Exhibit-Dev-The-Art-of-Writing-110.pdf (accessed on 12 December 2020).
9. Smithsonian Institution. A Guide to Exhibit Development. Technical Report. Available online: https://share.google/uNpVuDmoa7Z042NXC (accessed on 5 April 2018).
10. Falk, J.H.; Dierking, L.D. *The Museum Experience Revisited*; Routledge: Oxfordshire, UK, 2016.
11. International Council of Museums. ICOM Code of Ethics for Museums. Technical Report. International Council of Museums. Available online: https://icom.museum/wp-content/uploads/2018/07/ICOM-code-En-web.pdf (accessed on 20 July 2017).
12. Zheng, C.; Yu, Z.; Jiang, Y.; et al. ArtMentor: AI-Assisted Evaluation of Artworks to Explore Multimodal Large Language Models Capabilities. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Yokohama Japan, 26 April 2025–1 May 2025; pp. 1–18.
13. Chakrabarty, T.; Laban, P.; Wu, C.S. Can AI Writing be Salvaged? Mitigating Idiosyncrasies and Improving Human-AI Alignment in the Writing Process Through Edits. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Yokohama Japan, 26 April 2025–1 May 2025; pp. 1–33.
14. Laban, P.; Vig, J.; Hearst, M.; et al. Beyond the Chat: Executable and Verifiable Text-Editing with LLMS. In Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology, Pittsburgh, PA, USA, 13–16 October 2024; pp. 1–23.
15. Bin, Y.; Shi, W.; Ding, Y.; et al. Gallerygpt: Analyzing Paintings with Large Multimodal Models. In Proceedings of the 32nd ACM International Conference on Multimedia, Melbourne, Australia, 28 October–1 November 2024; pp. 7734–7743.
16. Dhillon, P.S.; Molaei, S.; Li, J.; et al. Shaping Human-AI collaboration: Varied Scaffolding Levels in Co-Writing with Language Models. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024; pp. 1–18.
17. Mirowski, P.; Mathewson, K.W.; Pittman, J.; et al. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–28 April 2023; pp. 1–34.
18. Anderson, B.R.; Shah, J.H.; Kreminski, M. Homogenization Effects of Large Language Models on Human Creative Ideation. In Proceedings of the 16th Conference on Creativity & Cognition, Chicago, IL, USA, 23–26 June 2024; pp. 413–425.
19. Lee, M.; Gero, K.I.; Chung, J.J.Y.; et al. A design space for intelligent and interactive writing assistants. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024; pp. 1–35.
20. Song, G. Cross-cultural exchange of heritage in museums: A study of Dunhuang art in the Hong Kong Heritage Museum from the perspective of cultural translation. *Crit. Arts* **2023**, *37*, 85–99.
21. Guillot, M.N. Cross-Cultural Pragmatics and Translation: The case of Museum Texts as Interlingual Representation. In *Translation: A Multidisciplinary Approach*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 73–95.
22. Sakib, S.K.; Das, A.B. Challenging fairness: A Comprehensive Exploration of Bias in LLM-Based Recommendations. In Proceedings of the 2024 IEEE International Conference on Big Data (BigData), Washington, DC, USA, 15–18 December 2024; pp. 1585–1592.
23. Tao, Y.; Viberg, O.; Baker, R.S.; et al. Cultural bias and cultural alignment of large language models. *PNAS Nexus* **2024**, *3*, gae346.
24. Jiang, N. The "Outline, Subject, Form" Curatorial Trilogy—Taking the Robot-themed Exhibition as an Example. Available online: https://mp.weixin.qq.com/s/296ahrrAzLEzVb_jgcjz8g (accessed on 28 March 2024).
25. Ravelli, L.J. Making language accessible: Successful text writing for museum visitors. *Linguist. Educ.* **1996**, *8*, 367–387.
26. Perego, E. Into the language of museum audio descriptions: A corpus-based study. *Perspectives* **2019**, *27*, 333–349.
27. Serrell, B. *Judging Exhibitions: A Framework for Assessing Excellence*; Routledge: Oxfordshire, UK, 2017.

28. OpenAI. GPT-4o. Technical Report. Available online: https://openai.com/index/gpt-4o/ (accessed on 8 August 2024).
29. Comanici, G.; Bieber, E.; Schaekermann, M.; et al. Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities. *arXiv* **2025**, arXiv:2507.06261.
30. Liu, A.; Feng, B.; Wang, B.; et al. DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model. *arXiv* **2024**, arXiv:2405.04434.
31. Pan, S.; She, J. Tanka Heritage Revived: AI-Generated Artworks in Three Chinese Art Styles. In Proceedings of the 17th International Symposium on Visual Information Communication and Interaction, Hsinchu, Taiwan, 11–13 December 2024; pp. 1–2.
32. Wang, Y.A.; She, J.; Lin, T.T.; et al. AI Visual Art History: An Art Movement with Expanded Artistic Horizon. *ACM Trans. Multimed. Comput. Commun. Appl.* **2025**, *21*, 1–16.