

MultiVerse: A Creator-Centered Approach to Steering Context-Adaptive Lyrics

Alexander Wang
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Chris Donahue
Computer Science Department
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

David Lindlbauer
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

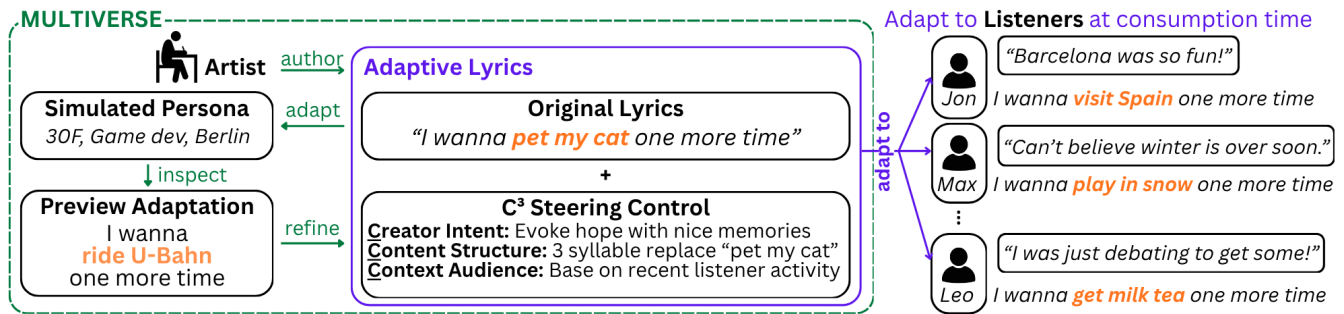


Figure 1: We anticipate the rise of AI-driven trends in music and multimedia like *adaptive lyrics*, where lyrics are dynamically tailored to different audiences at consumption time. Because final outputs are generated at consumption time, artists cannot directly review or edit each variation, running the risk of deviating from their creative intent. MultiVerse addresses this by providing tools that allow artists to iteratively author steering controls over intent, content structure, and audience context, while previewing how lyrics may adapt across audiences with simulated personas. Once the controls are refined, they can be deployed to deliver adaptive lyrics to real listeners at scale.

Abstract

Generative AI may enable new forms of context-aware creative expression by dynamically tailoring media content to its consumption context. For instance, AI systems could adapt song lyrics to the listener and their current activity. However, existing media adaptation systems primarily optimize for audience experience, often neglecting artists' intent, style, and preference. We address this challenge by introducing a novel creator-centered approach to adaptive media authoring and present MultiVerse, a system that instantiates this approach for steering adaptive lyrics. Our approach allows creators to explicitly author controls based on their intent, lyric structure, and audience context, and uses rule-based validations to ensure controls are followed. We conducted a study with 10 songwriters, comparing MultiVerse with a prompting-based workflow for composing adaptive lyrics. The comparison revealed that creators preferred to author how lyrics adapt by explicitly specifying relevant context and adaptation constraints, while recognizing trade-offs in flexibility and iteration speed. Interviews further revealed that creators viewed adaptive media as enabling new forms of audience connection, introducing a distinct creative process, favoring new compositional strategies, and reshaping notions of authorship.

CCS Concepts

• **Human-centered computing** → *Interactive systems and tools; Mixed / augmented reality*; • **Applied computing** → *Media arts*.

Keywords

Adaptive Media; Authoring Tool; Music

ACM Reference Format:

Alexander Wang, Chris Donahue, and David Lindlbauer. 2026. MultiVerse: A Creator-Centered Approach to Steering Context-Adaptive Lyrics. In *The 39th Annual ACM Symposium on User Interface Software and Technology (UIST '26)*, November 02–05, 2026, Detroit, MI, USA. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3830398.3830530>

1 Introduction

Consider the following example: A listener is boarding a late-night flight to start a new job abroad, leaving behind their hometown and family. As they sit at the gate listening to *Don't Stop Believin'* by Journey, the original verse ("Just a city boy, born and raised in South Detroit. He took the midnight train goin' anywhere") *adapts* to reflect their situation: "Just a quiet girl, leaving home in search of more. She took a red-eye flight bound for San Jose." The modified lyric mirrors the listener's circumstances, making the song feel more personally relevant—as if the song briefly breaks the fourth wall, allowing the artist to directly address the listener's own moment of transition.

Adaptive lyrics, like in the earlier example, demonstrate how songs can respond to individual listeners to create a more personal experience. In practice, artists have long adapted their work to



This work is licensed under a Creative Commons Attribution 4.0 International License. *UIST '26, Detroit, MI, USA*

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2856-3/2026/11
<https://doi.org/10.1145/3830398.3830530>

engage different audiences. At a broad scale, they create region-specific editions with translated or culturally localized content (e.g., Nena’s *99 Luftballons* and *99 Red Balloons*), and clean radio edits that tone down profanity and violence. At a finer scale, they often tailor certain lines during live shows to reference shared experiences with the audience. However, this quickly becomes impractical to scale across the countless contexts in which listeners encounter music.

Recent advances in AI, however, are breaking through that barrier—enabling music to be automatically adapted in real-time based on the listening context. Early work in this space has primarily focused on listener-side benefits, such as reducing disruption [75], improving awareness [39], and boosting exercise motivation [76]. Notably, Wang et al. introduced a method that modifies song lyrics to reflect real-time notifications [77]—e.g., replacing original lyrics with “your flight to NY is boarding now.” Adaptive lyrics could thus adjust automatically to individual audience members, e.g., their preference, location, or aspects of their personality. This personalization can increase the connection between artists and their audience, and improve the experience for the audience [42]. This approach, however, comes at the cost of taking control from artists, and potentially deviating from their intent. It is thus unclear, how artists can retain control over AI adaptations of their original work.

Motivated by prior formative work on creator perspectives [42], we highlight two main challenges of adaptive media. First, artists need ways to *control* how their creations are adapted to audience members. This includes specifying which parts could be modified and which parts must remain untouched, or controlling how aspects of an audience member should be infused into the medium. Second, since adaptive lyrics are generated at consumption time, artists cannot fully anticipate how each audience member will experience their work. Artists thus need ways to systematically preview how their original work is adapted for audience members, and take this into account for steering the adaptive media.

To address this, we propose C^3 , a new creator-centered paradigm for adaptive media authoring, where artists explicitly steer how their work responds to consumption context rather than fully delegating adaptation to automated systems (Figure 1). Our approach is grounded in three design axioms corresponding to three factors in adaptive media: creator, content, and context. Rather than handling these factors implicitly within automated generation pipelines, our approach exposes them as explicit dimensions for steering. Creators specify how their work aligns with their intent (creator), define structural constraints that must be preserved across variations (content), and articulate which consumption-side conditions are relevant and how adaptation unfolds in response to them (context).

To operationalize C^3 and examine its implications in practice, we instantiated this approach in the domain of songwriting through the development of a tool for steering adaptive lyric content. We contribute MultiVerse, an adaptive lyric steering system that exposes creator, content, and context as structured controls for lyric adaptation. MultiVerse explicitly exposes context as simulated personas for authors to preview how lyrics adapt to different audiences, how they react to the lyrics, and allow creators to proactively query for specific audience information to shape adaptation behavior. MultiVerse implements domain-specific tools to control content structure such as annotating the desired stress pattern, syllable

count, and which words should rhyme. The system further incorporates rule-based validation mechanisms that check generated lyrics against musical and structural requirements, ensuring coherence while constraining the space of possible variations.

Envisioned usage. We envision that artists leverage MultiVerse to adapt their *existing lyrics*, as outlined in the walkthrough (Sec. 2). Given original lyrics, they author *steering controls* that guide how lyrics can be adapted for different audience members. MultiVerse allows artists to preview adaptations for simulated personas and see how those personas might respond to the lyrics. The *output* of MultiVerse is a set of *steering controls*, defined for use in conjunction with the original lyrics. When audience members listen to the songs, a future system (e.g., a plugin for Spotify or Apple Music) would adjust the lyrics based on these *steering controls*. Contextual information could be provided directly by listeners (e.g., demographics, occupation, favorite food) or extracted automatically by other systems (e.g., location). This would allow listeners to experience personalized versions of songs, such as hearing their favorite artists sing about the city they live in. In this paper, we focus on exploring how artists maintain control over their work, leaving the study of audience perceptions and preferences for future work.

To understand how C^3 and MultiVerse intersect with creative practice, we contribute insights from a study with 10 songwriters. Participants transformed their own compositions into adaptive versions using two contrasting workflows: MultiVerse, and a prompt-based LLM workflow representative of contemporary generative lyric practices.

Our study provides two complementary sets of insights: first, comparing MultiVerse with prompting revealed that creators valued mechanisms for defining meaningful adaptation context and explicit control over adaptations, while recognizing trade-offs in adaptation flexibility and iteration speed; second, interviews revealed broader creator perspectives on adaptive media, showing how it enables new forms of audience connection, introduces distinct creative processes and compositional strategies, and reshapes notions of collaboration, authorship, and ownership.

In summary, our work contributes:

- (1) C^3 , a **creator-centered authoring approach for adaptive media** that exposes creator, content, and context as explicit dimensions in the authoring workflow.
- (2) **MultiVerse, an adaptive lyric steering system** that operationalizes C^3 through structured controls and rule-based validation of adaptations.
- (3) **A study with songwriters (N = 10)**, providing empirical insights into how creators engage with C^3 and the trade-offs between explicit structured control and prompt-based workflows.

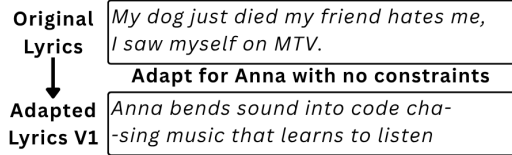
2 Walkthrough

To better illustrate C^3 and its instantiation in MultiVerse for adaptive lyrics, we provide a walkthrough of a creator iteratively authoring steering controls. Suppose a creator is using MultiVerse to author steering controls. Their original lyrics are “*My dog just died my friend hates me, I saw myself on MTV*” and their goal is to have them adapt dynamically to each listener’s context — while respecting a few key constraints: (i) maintaining the original rhyme scheme, (ii) preserving the theme of isolation and feeling lost, and

(iii) connecting to audience contexts that resonate with that theme. They begin by focusing on a single example user persona, then validate the pipeline across a broader set of audiences.

No constraint (Figure 2). The creator starts by adapting lyrics with no controls for *Anna*, a young graduate student studying music and CS, they observe the lyric: "*Anna bends sound into code, chasing music that learns to listen*," which shifts focus entirely to the persona, producing a line that no longer reflects the artist's original intent.

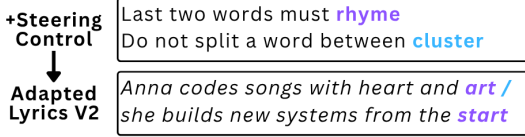
Figure 2: Adapt for persona Anna with no constraints.



Content structure (Figure 3). In the original melody, there is a long pause after the word "me" (1 syllable), which is currently being replaced with "chasing", a 2 syllable word that now splits across two clusters of rhythm. Furthermore, "me" and "V" in the original lyrics rhyme. To fix this, the creator uses the rhyme tool to bind note 8 with note 16, and the cluster tool to split note 8 with 9.

Using the updated control settings, the new output becomes: "*Anna codes songs with heart and art, she builds new systems from the start*." The lyrics now fit the original composition, but still diverge from the theme of the song.

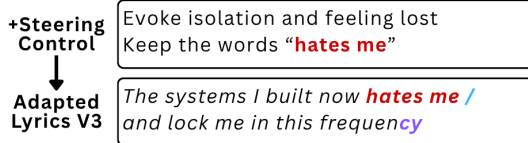
Figure 3: Artist specifies content structure.



Creator intent (Figure 4). When initially composing the song, the creator intended the music to convey themes of isolation and feeling lost, and this specific line should convey the feeling of being rejected by those around them. To fix this, the creator uses global intent to specify the song's theme and the word lock to ensure that the output keeps the exact words "hates me."

Using the updated control settings, the new output becomes: "*The system I built now hates me, and lock me in this frequency*." The lyrics now fit the overall theme of the song, but still feels artificial and only adapt based on superficial aspects of the listener.

Figure 4: Artist specifies creator intent.

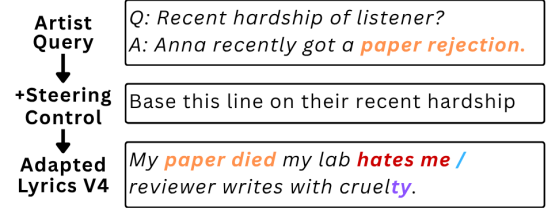


Audience context (Figure 5). To effectively tailor the content to Anna, the virtual persona, the creator uses the contextual query tool to proactively ask specific questions about the listener

and reveal key contexts relevant to the music. The creator asks, "what happened recently that makes you feel lost?", and MultiVerse responds that "A recent hardship for Anna is having their paper rejected. The harsh comments from reviewers and isolation from their research lab leave them feeling lost."

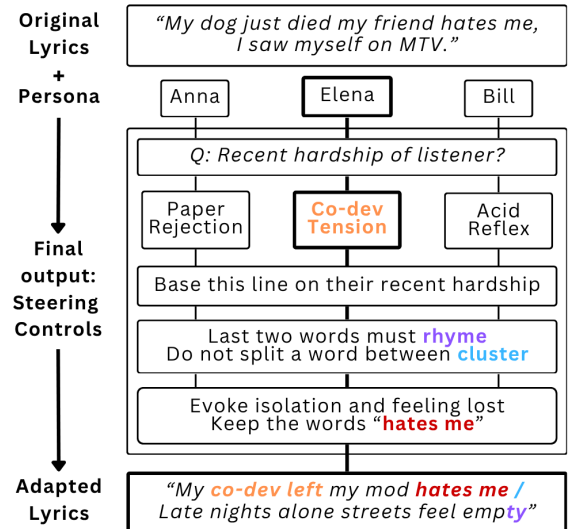
With the updated audience context, the new output becomes: "*My paper died my lab hates me, reviewer writes with cruelty*." The creator is now satisfied with this output.

Figure 5: Artist specifies relevant audience context.



Generalizing across audiences (Figure 6). While the lyrics for Anna are satisfactory, the creator next evaluates how well the same set of steering controls generalize across several virtual personas. For example, Elena is an indie game designer in Berlin who collaborates closely with online creative communities and helps run a Discord server for her game. Through contextual query, the system simulates how the authored controls would adapt the lyrics given her personal context—such as tension with a collaborator—while preserving structural and thematic constraints. The creator repeats this process for multiple personas, making minor adjustments whenever unexpected adaptations arise. They confirm that the steering controls produce adaptations that are well aligned with their intent, and finalize them for application to the original lyrics, enabling the song to dynamically adapt for real listeners at scale.

Figure 6: Artist preview output for other personas.



3 Related Work

3.1 Generative AI for media

Generative AI is now capable of producing multimedia with high fidelity and controllability, and increasingly allows fine-grained edits to existing media content, enabling systems to dynamically adapt media during consumption. Notably, commercial tools such as Google Gemini (Nano Banana [25]) allow users to freely manipulate images via text prompts, enabling operations such as adding or removing objects, altering character poses, and transforming visual style with fine-grained control.

Similarly, recent models in the music domain also demonstrate notable advances in fine-grained control [83, 84], real-time inference [6, 17], content infilling [34, 73], and editing [74]. For instance, ACE-step enables modifying input music to change the singer, style, and lyrics [24]. While these capabilities are still emerging, they signal a broader shift: the kind of open-ended manipulation already possible in images is soon coming to music and other creative domains. This rapid rise of malleable content creates a pressing need for frameworks that give creators control over media adaptation.

3.2 Context-aware interfaces and media

Computing systems are increasingly able to sense and respond to the context in which they operate [15], enabling digital experiences that adapt to the surrounding physical environment and user state and actions [38, 40, 62]. While substantial prior work focuses on adapting interface elements without altering the underlying content [9, 12, 18, 31], context-awareness is applied to modify the content itself, shaping storytelling media across modalities such as VR [72], film [32], video captions [35], books [28], journalism [2], and scientific communication [3, 41].

Similar context-aware adaptations have also been explored in music. Early work primarily focused on recommending songs suited to contextual factors such as mood, location, cultural background, and conversations [8, 43, 66, 87, 88]. More recent work adapts the music itself, including structure-level modifications that align musical transitions with activities such as driving or weightlifting [39, 76], as well as content-level changes that alter existing song elements, such as integrating ringtones as new melodies or replacing lyrics with speech from conversational agents [75, 77]. Currently, prior work in context-adaptive systems either adapts interface elements without considering the underlying content, or modifies content automatically without regarding creator’s intent (Figure 7).

3.3 Tools for authoring static media

Generative AI has increasingly been explored as a creativity support tool for producing many forms of static media content, including text [7, 22, 49, 71], image [5, 13, 51], and video [36, 53, 78, 79, 85]. In music, AI-assisted systems support artists in iteratively exploring ideas, refining outputs, and collaborating with generative models, accelerating ideation while raising questions around creative agency and authorship [20, 59]. Prior work introduces mechanisms for steering generation [23, 54, 55], incorporating multimodal inspirations and queries [4, 44], and expressing abstract moods to guide outputs [30]. Generative approaches have also been

applied to songwriting, including generating lyrics under structural constraints and composing melodies aligned with existing lyrics [1, 16, 48, 70, 77, 81, 82].

While prior work focuses on improving creator control over static generated outputs, our setting departs from this assumption: the final output is not fixed, but dynamically generated at consumption time. As a result, creators cannot directly inspect or anticipate all possible variations, necessitating new approaches to encoding and preserving artistic intent.

3.4 Tools for authoring dynamic media

Currently, the most mature ecosystem for creator-authored dynamic media exists in interactive content, such as video games and interactive narrative systems that allow players to progress through stories non-linearly. These systems typically adopt branching narrative structures, in which player choices route them along pre-authored paths toward distinct narrative outcomes [64]. Authoring tools for such systems often rely on node-based environments (e.g., Twine, Ink, ChoiceScript), enabling creators to script conditional logic and manage narrative branches.

Prior work has further explored systems that generate narrative content and gameplay behavior at runtime [19, 21, 37], moving beyond fixed branching structures. These approaches demonstrate how generative models can alter storylines dynamically based on player interaction, enabling personalization that is no longer constrained to a small set of predefined paths. This shift enables a form of authoring centered on world planning [14, 47], defining individual agent behavior [60, 80], and constraining the space of possible narrative trajectories [56, 65], rather than manually scripting every plot branch and dialogue line.

Beyond interactive media, recent research began to explore authoring other forms of dynamic media that adapt to audience context as opposed to explicit interaction. Kim et al. investigate creators’ perspectives on AI-mediated personalization in creative writing through interviews with authors considering hypothetical scenarios of AI-assisted narrative adaptation [42]. While participants recognized the potential for adaptive media to increase audience resonance, they emphasized mechanisms for preserving authorial intent. Their findings highlight the importance of developing new authoring tools for adaptive media beyond interactive narratives—the primary focus of prior work—to other emerging forms of dynamic media such as adaptive music.

4 C^3 for Adaptive Media

We propose the C^3 framework to support creators in authoring adaptive media.

Unlike static media, adaptive media is not authored as a fixed artifact, but as behavior that unfolds across future consumption contexts. This changes the role of the creator from editing individual outputs to specifying how media should adapt while remaining faithful to artistic intent. C^3 addresses this challenge by externalizing three core elements of adaptive authoring—creator, content, and context—as explicit representations that creators can inspect, author, and refine, as illustrated in Figure 7.

Defining adaptive media. We refer to adaptive media as content that responds to characteristics of the consumption context, drawing parallels from context-aware and ubiquitous computing [15, 27, 50]. Adaptive media could change based on the audience—their background, preferences, or current state; the environment in which it is consumed; tasks or activities performed by the audience during consumption; or the medium through which it is delivered. This poses unprecedented challenges for creators, emphasizing the need for frameworks that enable meaningful artist control.

Authoring adaptive media. Unlike static content, creators cannot directly review every possible outcome, and unlike interactive media, adaptive media responds implicitly to contextual factors, making it difficult for creators to predict or fully conceptualize all possible variations. In interactive systems such as video games and narrative experiences, variation is driven by *explicit* user inputs, allowing authors to define a bounded space of possible trajectories within the work [56]. In contrast, adaptive media adjusts *implicitly* based on those contextual parameters as either a complementary or additional factor. Contextual factors may differ substantially from the author’s own perspective and lived experiences. Since creators need to define and consider context as additional input, we argue that this has implications for the whole workflow of creation.

A framework for adaptive media. Combining insights from interactive media authoring [26, 56, 64] and context-aware computing [15, 27, 50], we define three design axioms (DAs) that capture the core characteristics of systems developed under our C^3 framework.

- **DA1: Consider creator intent.** Provide mechanisms for creators to articulate artistic goals to guide adaptations.
- **DA2: Consider content structure.** Provide mechanisms for specifying which structural aspects must be preserved during adaptation.
- **DA3: Consider context-driven adaptations.** Provide mechanisms for shaping relevant audience context, and explore how media adapt for different contexts.

Following these goals, we conceptualize adaptive media authoring as the coordination of these three interdependent dimensions: *creator*, *content*, and *context*. Rather than treating adaptation as a purely automated generation process, C^3 exposes these dimensions as explicit elements of the authoring workflow. This allows creators

to specify how adaptations should align with their artistic goals, structural constraints, and contextual signals. For each dimension, we propose to consider a set of domain-agnostic inputs and steering actions. We also provide domain-specific instantiations for the context of adaptive lyrics, which we implement in MultiVerse. We see our approach as extensible within the individual categories, since each domain has specific considerations.

Creator. Adaptive media authoring considers the *artistic intent* that guides how media adapts. As *inputs*, systems should consider parameters that capture the envisioned theme, narration voice and style, emotional message, and artistic vision. Systems should *steer and constrain* the space of how media is adaptive by considering *where* those parameters are applied, for example globally enforcing a theme; *how* those parameters are applied, for example how narration voice influences the media; which of those aspects are *prioritized*, for example emphasizing narrative voice locally over a global theme to preserve a certain style. Finally, we define *boundaries* of the adaptations, for example how far an adaptation can deviate from the original lyrics and the defined intent.

Creator considerations in adaptive lyrics. Lyrics are semantically dense, where small changes can shift meaning, tone, or style. MultiVerse lets creators steer adaptations at three levels: global anchors (theme, emotion, style), local line guidance, and word-level locks. This explicit separation ensures AI-generated variations remain aligned with creator intent while allowing controlled flexibility.

Content. Content captures the structural properties of a media artifact. Different creative domains impose different structural requirements. For example, structural properties such as camera framing and the sequence of events are critical for narrative coherence in film editing [46].

In general, authoring tools need to enable creators to specify what aspects to *preserve*, what parts can be *relaxed/adapted*, and finally what can be *added* or *removed*. Those actions can be defined on a global level (e. g., whole characters can be added to a story) and tools need to enable creators to address downstream consequences. Alternatively, content can be steered locally, e. g., word- or sentence-level edits for text-based media, or color adjustments for image-based media. Importantly, not all structural properties are equally important for every work. Creators therefore need to be able to specify which structural elements must be preserved and how adaptations should respect these constraints.

Content considerations in adaptive lyrics. Lyrics consist of a rich set of structural properties, which have been explored in musicology [58] and music information retrieval [57]. MultiVerse enables creators to steer adaptive lyrics by constraining factors such as syllable count, rhyme structure, stress patterns, and melodic alignment, ensuring that generated variations remain compatible with the underlying melody and overall song structure.

Context. Context is the situational information that drives the adaptation of media artifacts, including aspects about end users (or audience), environment and system capabilities [50]. Each dimension has multiple parameters, such as an audience member’s preferences or prior experiences. Adaptive authoring tools should allow creators to *define* which contextual inputs are considered,

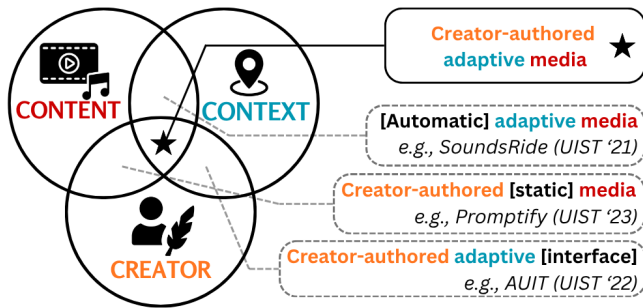


Figure 7: C^3 unifies creator, content, and context to support creator-driven adaptive media.

specify their *influence*, determine their *priority*, and *constrain* how contextual information guides adaptations.

Context considerations in adaptive lyrics. Audience and situational context shape how lyrics are interpreted and experienced. While a song could, in principle, adapt to many contextual signals at the time of consumption (e. g., listening location), in this work we focus on the audience as the primary driver of adaptation. MultiVerse enables the prototyping of this process through simulated listeners. Creators can proactively define which audience aspects are relevant, and then guide how adaptations are applied by linking those responses to specific parts of the lyrics.

5 MultiVerse

To explore C^3 in practice, we instantiate our approach in the domain of songwriting through the design of MultiVerse, a system that enables creators to steer adaptive lyrics by authoring steering controls and testing adaptation behavior with simulated personas.

Concretely, MultiVerse takes a finished set of original lyrics as *input*. Creators then author steering controls that specify how the lyrics should adapt across different audience contexts. Using simulated personas, creators preview potential adaptations to explore variations and iteratively refine their controls. The *output* of this process is a set of steering controls that guide how the lyrics adapt. In this section, we describe the design and implementation of MultiVerse, organized around the design goals: features that support the specification of creator intent (5.1), content structure (5.2), and adaptation context (5.3).

5.1 Creator: authoring artistic intent

MultiVerse provides multi-level control for steering adaptive content, supporting both high-level guidance of intent and low-level control over line-level adaptations (Fig 8).

Global intent anchors. MultiVerse supports the alignment of high-level artistic intent through specifying a set of global anchors based on 3 dimensions:

- Theme: the core message and concept of the song.
- Emotion: the target emotion to evoke.
- Style: the voice and aesthetics of the lyrics agnostic to theme and emotion.

Global anchor
Theme Anchor

Explore personal hardship and the sense of feeling of lost.

Emotion Anchor
Style Anchor

Disorientation

Alienation

Self-deprecating

Local anchor
Word lock
Do not change

This line should focus on isolation/rejection

Mydogjustdiedmyfriendshateme

IsawmyselfonMTV

Example Output: “The systems I built now **hate me**, and lock me in this frequency.”

Figure 8: Creator intent authoring features in MultiVerse. Global anchors affect all generations, local anchors affect one line, and word locks keep specific words unchanged.

^ Accent
Rhyme
Cluster

My dog just died my friends hate me
 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1

Rhyme “art”/“start”
 Cluster of (8/8)
 Accent at note 4

I saw myself on MTV
 1 1 0/1 1 1/1/1
 1 1 2 1 3

Syllables: 16 (Clusters: 8, 8)

Example Output: “Anna codes **songs** with heart and **art** / she builds new systems from the **start**.”

Figure 9: Content structure authoring features in MultiVerse. Accent ensures the labeled syllable is stressed, Rhyme ensures selected words rhyme, and Cluster prevents multisyllabic words from crossing labeled boundaries.

The lyrics are automatically analyzed to generate an initial set of tags, while also supporting iterative refinement of this document with the creator. This anchor document is provided to the prompt for all lyric generation tasks to steer the output. Emotion and style anchors are short keyword tags, while the theme anchor is a more elaborate statement in natural language.

Local intent anchors. MultiVerse additionally supports the optional authoring of short intent statements to steer the behavior of each individual line.

Word locks. MultiVerse allows creators to lock or unlock individual words, ensuring they remain unchanged during adaptive generation.

5.2 Content: authoring music structure

MultiVerse provides tools for controlling structure (Figure 9).

Syllable count. When writing lyrics to a predetermined melody, syllable count must match the number of vocal notes, as each note typically carries one syllable. MultiVerse enforces this as a hard constraint to preserve alignment with the fixed melodic structure.

Melody clustering. Even when the line-level syllable count is correct, word-level coherence can still be disrupted if a word is split across discrete melodic segments. For example, when the lyrics “misleading list” is sung with a pause in the middle, it may sound like “miss Lee (pause) dean list” and confuse the listener. To prevent this, MultiVerse allows creators to label where breaks occur. The system then ensures that the number of syllables in each cluster matches the corresponding notes.

Stress pattern. Beyond syllable count, natural-sounding lyrics require alignment between lyrical stress and melodic emphasis. In many melodies, certain notes fall on strong beats or receive emphasis through duration or pitch. Placing an unstressed syllable in these positions can sound awkward or alter the meaning. For example, the word “present” can be pronounced as *PRE-sent* (a gift) or *pre-SENT* (to demonstrate). To prevent this, MultiVerse allows creators to mark syllable positions that should be lexically stressed.

1 **B** Bill, 45M, Truck Driver, Rural Midwest

"Bill, 45, is a long-haul truck driver who measures his life in miles and diesel prices, finding meditative peace in the 11-hour stretches behind the wheel on Midwestern highways."

2 **It sounds like** **B** "My kids probably know what that is, but I don't."

iPod Touch yellow Pikachu case

3 **CHILDHOOD DEVICES** what are childhood devices that you owned?

Childhood devices that Bill owned include a wooden Zenith console television

4 **It sounds like** **[Childhood Devices]** **iPod Touch** **yellow Pikachu case**

5 **It sounds like** **B** "Reminds me of watching the weather with my old man."

Old Zenith tubes warming in the night

Figure 10: Persona-based authoring in MultiVerse. Creators query listener context, bind responses to lyric variables, and preview personalized lyric adaptations.

Rhyme. Rhyme plays a key role in lyrical structure and listener expectation. MultiVerse allows users to label pairs of words as rhyming constraints, ensuring that any subsequently generated lyric variation preserves the specified rhyme relationships.

5.3 Context: Audience-aware authoring

MultiVerse enables audience-aware authoring through simulated personas, allowing creators to preview feedback and define relevant audience context to shape system behavior (Figure 10).

Listener personas. MultiVerse includes a set of fictional listener personas that represent diverse audience contexts. Each persona is described through nine profile dimensions capturing identity, background, and lifestyle factors that may shape how a listener interprets a song. These dimensions include name, demographics, location and cultural context, occupation, background and formative experiences, social relations, physical appearance or aesthetic style, personality traits, and daily routines.

Contextual query and variables. While personas provide a rich contextual profile, they cannot capture every aspect of a listener's experiences. In principle, an audience member's background could be described through an arbitrarily detailed biography, which would be impractical to represent within the system or incorporate into generation prompts.

To address this limitation, MultiVerse allows creators to proactively define the aspects of audience context that are most relevant to their song through *contextual queries*. A contextual query is a question posed to the listener persona that elicits information aligned with the song's themes or narrative. The response can then be linked to *variables* that replace specific text in the lyrics.

Listener feedback. The system simulates fictional audience personas, showing their contexts and reactions to different lyrics. By observing these simulated responses, creators can adjust controls to better align adaptations with intended audience experiences.

5.4 Implementation

The interface of MultiVerse is implemented as a web application using React. Adaptive lyric generation is performed through API calls to Gemini 3 Flash, and then validated with rule-based computational linguistic library RiTa.js [33].

Preprocessing and control initialization. When a song is loaded, MultiVerse initializes creator controls by using RiTa.js to infer structural properties (e.g., syllable count, stress, and rhyme). These automatically populated controls serve as editable defaults that creators can review and override. The system also derives an initial creative vision (theme, emotion, and style) from the source lyrics.

Adaptive lyric generation. To generate lyrics, MultiVerse prompts Gemini using user-authored controls together with persona information (Fig 11). Because the generated lyrics must satisfy strict structural constraints (e.g., syllable count, rhyme bindings), generation is performed on a per-line basis, with the prior line available in context. This allows the system to validate each candidate line deterministically before proceeding. The prompt is constructed from modular components representing creator-defined constraints, such as locked words, variable spans, rhyme bindings, intended line meaning, listener context, and song-level creative goals.

Batch generation. Rather than generating a single output line at a time, the system requests 15 candidate lyric lines in each attempt. This provides a pool of alternatives that can be filtered and ranked.

Prompt variation. The system also cycles through several prompt formulations. The default prompt asks the model to refer to the structure of the original line, while a second prompt asks the model to ignore the original line. A third prompt provides a partially complete template and asks the model to complete the lyric using the provided fragments. If a line contains locked words or variables, the template-completion prompt is prioritized.

Constraint validation. Each batch of generated candidates is filtered using deterministic validation checks. For verifiable constraints—including syllable counts, phrase-cluster boundaries, locked-word preservation, accent placement, and rhyme matching—MultiVerse applies rule-based checks implemented with RiTa.js. If no candidate in a batch satisfies the constraints, the system proceeds to the next prompt variation. Otherwise, the valid candidates from the first successful batch are retained for further evaluation.

Candidate ranking. After validation, the remaining candidates are evaluated with a second LLM call. The model assesses them from the listener's perspective and ranks them by expected preference, and the top-ranked candidate is selected as the final output.

6 User Study

We conducted a user study with songwriters to explore how creators approach the task of steering adaptive lyrics and how different workflows shape this process. As there is no established method for this task, our goal was not to evaluate performance against a canonical approach, but to use contrasting workflows to surface trade-offs in how this emerging form of media can be steered. We examine how creators reason about encoding adaptation into their

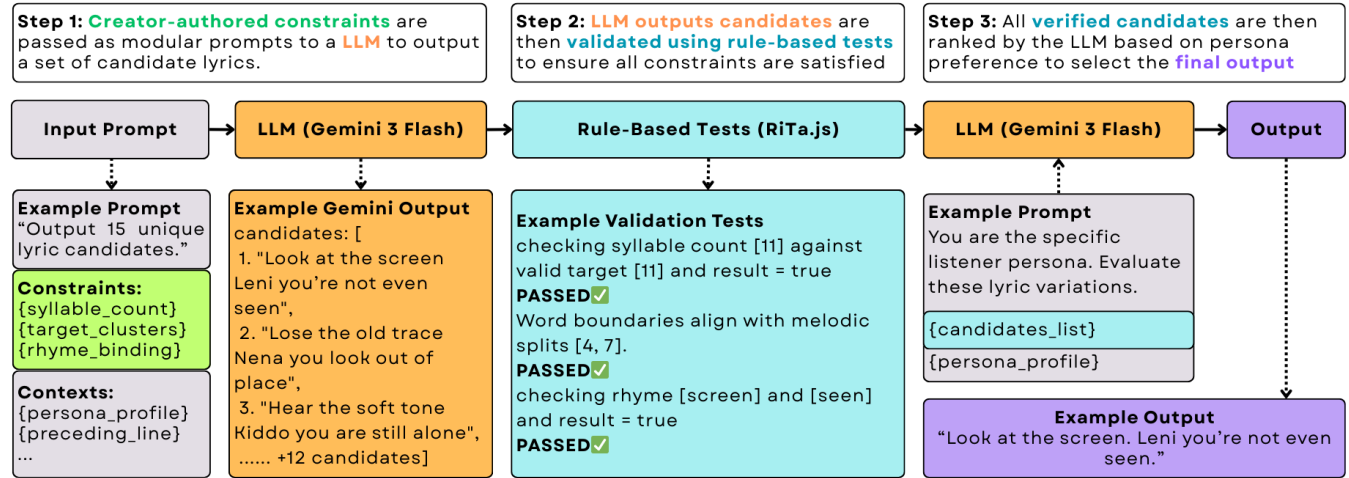


Figure 11: Lyric generation pipeline: Creator-authored constraints are passed as modular prompts to the LLM to generate an initial set of candidates. Each candidate is then validated using rule-based tests to ensure all constraints are satisfied. All verified candidates that pass the tests are then ranked by the LLM based on persona preference to select the final output.

songs, what trade-offs arise, and how these workflows affect their sense of control and authorship.

Workflows. Participants used their own songs to create adaptive lyrics using two contrasting workflows: **MultiVerse**, and **Prompt**, a prompt-based workflow that relies on free-form text instructions. As there is no established method for steering adaptive lyrics, the prompt interface represents a hypothetical approach that is similar to everyday AI chat tools. Relative to the structured approach of MultiVerse, Prompt is more “hands off”—all creator steering is conveyed through a text box and executed by an LLM (Gemini). The prompt-based workflow features an interface consisting of a persona selector identical to full MultiVerse, windows to view original lyrics and generated lyrics, and a text box to input the adaptation instructions. Unlike AI chatbots, Prompt does not allow back-and-forth conversation, as no artist iteration is allowed during runtime adaptation. Instead, both workflows involve a prompt template requesting Gemini to alter {original lyrics} based on {intended audience} while adhering to {control instructions}. The adaptation instruction is applied to the entire set of lyrics rather than individual lines, and since the instructions are in free-form text, no rule-based validations can ensure they are followed.

6.1 Participants

We recruited 10 skilled participants (7 male, 3 female; age: $M = 23.7$, $SD = 3.47$) with at least 2 years of songwriting experience and at least 2 original compositions. Participants had diverse musical backgrounds, and represented both music students and professional musicians. On average, participants were highly skilled and reported $M = 9.2$ years ($SD = 6.14$) of songwriting experience and $M = 8.2$ years ($SD = 4.64$) of formal musical training. Several participants had professional experience composing, producing, or performing music, including releasing albums, composing for games and films,

performing live, and teaching music. Table 1 in the appendix provides details of participant backgrounds. Participants were recruited through online outreach and convenience sampling.

6.2 Procedure

Participants provided informed consent, demographic information, and were introduced to the study goals and the concept of AI adaptive lyrics. They then completed two conditions in counterbalanced order: MultiVerse, and Prompt. Each condition followed the same structure. Participants first completed a brief guided tutorial on the interface, including a walkthrough exercise for creating adaptive lyrics for a provided song (approximately 5 minutes for the prompt condition and 10–20 minutes for MultiVerse). They then performed a 20-minute authoring task in which they adapted one of their own songs (collected prior to the study and truncated to 8–16 lines) to support different audiences while preserving its original intent. In the MultiVerse condition, participants used the full interface to specify controls. In the Prompt condition, they authored instructions in natural language. In both cases, lyrics were generated using Gemini 3 Flash based on the original song, participant-authored controls, and target audience personas. All participants used the same 12 personas (four 3-persona sets). For each condition, one set of personas was used for authoring, and another set for evaluation, i. e., participants reviewed the outputs for a set of unseen personas to get a sense of how their adaptive lyrics would be used.

After each condition, participants completed a post-task questionnaire (7-point Likert scale) assessing their perceptions of the system and the resulting outputs. This includes 5 dimensions of the creativity support index [11] and 3 modified items from the mixed-initiative creativity support index [45].

After completing both conditions, they took part in a semi-structured interview reflecting on their experiences with the two workflows and broader perspectives on authoring adaptive music. Interviews were audio-recorded, transcribed, and analyzed through

affinity diagramming. We logged all participant-authored control interactions.

7 Results

We organize our results into two parts. We first present comparative findings from contrasting MultiVerse with prompting-based authoring, highlighting the strengths, limitations, and design trade-offs of different workflows. We then present broader creator perspectives on adaptive media, including why creators value adaptive media, how it reshapes the creative process, and how it influences notions of authorship.

7.1 Comparative insights

Overall, both workflows were positively received (e. g., “*I enjoyed using the system*” was rated $M = 5$ for both), with no statistically significant differences across subjective ratings (all $p > .05$, Table 2). Rather than indicating a clear preference for one workflow, participants described complementary strengths and trade-offs between the two approaches. Participants consistently appreciated the control and context exploration mechanisms afforded by MultiVerse, while also recognizing that the control introduced trade-offs in adaptation flexibility and iteration speed.

7.1.1 Context-based authoring mechanisms enabled creators to steer adaptations using meaningful context. Rather than adapting lyrics without regard for contextual relevance, participants wanted to specify contexts they considered meaningful for the song itself. As P4 explained, the system often by default “focuses on profession, but that’s not what I want to optimize for.” MultiVerse’s contextual querying mechanism allowed creators to proactively define which aspects of a listener’s context should drive adaptation. Nearly all participants (9/10) proactively queried contextual information and incorporated the information into their lyrics, most commonly asking questions aligned with their songs’ themes (6/10). For example, participants writing love songs asked about breakup details and regrets in love (P3, P5), while more introspective songs prompted questions about recent experiences relevant to the song’s emotional themes, such as recent regrets (P10). Participants also tailored queries to the structure of their lyrics (9/10), requesting contextual details that could be directly mapped into lyrical elements, such as sounds heard during work (P1), places where listeners reflect alone (P2), daily activities and environments (P8), current headspace and imagery (P9), or everyday habits such as where listeners get their news (P10). Participants described these creator-defined adaptation contexts as producing lyrics that felt more genuine and less “surface-level” than adaptations based only on generic audience attributes (P9).

7.1.2 Structured controls gave creators greater confidence over adaptation outcomes. Participants consistently associated MultiVerse with greater control over lyrics adaptation when compared to Prompt (8/10). Participants attributed this perception primarily to the ability to define explicit constraints on how lyrics could change, including guarantees that specific lines would remain unchanged. As P1 explained, MultiVerse “allows you to specify which parts stay original and which parts change, so it feels safer and more controlled.” In contrast, the simplified workflow was frequently

described as less consistent and predictable. Because instructions were embedded in free-form prompts, participants often had to “guess about how much of the prompt the AI would follow,” with the system sometimes ignoring parts of their instructions (P4). The resulting unpredictability also raised concerns about consistency across generations. As P10 explained, if the system were used in practice, they would be “nervous about generating the adaptation again and not liking the next version—even though I haven’t changed anything (in the control prompt).”

7.1.3 Strict constraints reduced adaptation flexibility and iteration speed. Participants observed trade-offs in lyric quality with strict constraint enforcement (7/10). While explicit constraints over accent, rhyme, and clusters improved structural consistency (P1), several participants felt they also narrowed the adaptation space. As P9 explained, constrained edits sometimes produced fragmented lyrics where “some of the lyrics were mine, some were generated, and some were edited,” whereas the prompt-based workflow could better fit a listener’s story, even if it diverged from the original intent. Participants also emphasized the importance of rapid iteration for refining adaptive lyrics (7/10). While the Prompt workflow supported rapid exploration through full-song generations (“the faster it is, the faster I can say whether I like it,” P5), MultiVerse’s more granular authoring workflow slowed iteration, leaving several participants feeling they had insufficient time to fully refine their controls during the study session.

7.2 Creator perspectives on adaptive media

Beyond comparing authoring workflows, participants reflected on adaptive lyrics as an emerging creative medium and how it differed from conventional songwriting. Our interviews revealed perspectives that extended beyond the specific interfaces, including why they found adaptive media valuable, how authoring adaptive media introduces a distinct creative process and new structural considerations for composition, and how it reshapes notions of collaboration, authorship, and ownership.

7.2.1 Creators valued adaptive media for enabling new forms of audience connection. Artists valued the personalization of adaptive lyrics for its potential to strengthen emotional connection with listeners. Study participants noted that adapting lyrics for specific listeners could make songs more meaningful and relatable, and help them achieve common goals of songwriting such as helping listeners “feel something” (P10) or providing experiences that are “really meaningful” and even “healing” (P4). Participants described adaptive lyrics as enabling fundamentally new relationships between artists and audiences. P9 envisioned that instead of searching for songs that happened to resonate with their lives, superfans could “customize lyrics to their experiences,” allowing their favorite artists to sing their stories. P4 similarly envisioned adaptive music becoming a new form of fan participation where fans co-create personalized tracks with their favorite artists, extending existing parasocial experiences such as K-pop personal chats that personalize artist-authored messages to create the illusion of one-to-one interaction [86]. P1 framed adaptive lyrics as a new form of lyrical “remixing,” where audience-specific details could “resonate so hard” by reflecting listeners’ own lived experiences.

Participants, however, also raised concerns about too much personalization. Tailoring lyrics too closely to a specific listener could reduce interpretive openness, as songs can “take on different meanings” (P6). Furthermore, lyrics may become “too hyper-tailored” to feel broadly relatable (P3), as well as lead to decreased perceived authenticity that feels “kind of fake” (P10).

7.2.2 Authoring adaptive media is a distinct creative process. Participants described adaptive lyric authoring as an iterative process of discovering which aspects of a song should remain invariant across future adaptations. Rather than treating artistic intent as something fully specified before authoring, they refined it through repeated cycles of generating adaptations, observing what changes challenged their intent, and updating their steering accordingly. As P6 explained, adaptive authoring required them to “deconstruct the lyrics,” and decide “which lines can be variable and which lines need to stay the same.” P7 similarly described the process as “a feedback loop” of “understanding what’s important in the lyrics” and “what about the lyrics is important to the artist,” then “using that information to generate more content” to continue the loop. Participants noted that this process also helped them better understand their own songwriting. P6 explained that this iterative process “made it very clear to me what the song was really about” and helped them “understand the mechanics of my lyrics much better.”

7.2.3 Adaptive media favored different content structures. Participants distinguished between songs that naturally lent themselves to adaptation and those that did not. P6 observed that some songs are “more adaptable than others,” explaining that lyrics referencing specific places naturally support variation across audiences. Similarly, P4 contrasted two of their own songs, noting that one built around abstract metaphors could benefit from adaptation because personalization might help listeners better understand its intended meaning, whereas another centered on seasons was already broadly relatable and required little personalization. In contrast, participants viewed deeply personal narratives as poor candidates for adaptation. As P9 explained, adapting a song about losing a parent could feel inappropriate because “it’s not your story to tell.”

This became particularly apparent when participants adapted finished songs. P10 explained that their songs were “already written” and “finalized in my head as the lyrics,” making it difficult to rethink how the song could vary. Similarly, P8 noted that their process would “look a lot different if I were starting from scratch.” Rather than composing freely, adaptation became “identifying certain phrases or words that might make sense to swap out” because the existing structure and rhyme scheme made the process “a lot more restricted.” Reflecting on the experience, both participants suggested they would compose adaptive songs differently from the outset, designing lyrical structures that better accommodate future variation.

7.2.4 Adaptive media reshaped notions of collaboration, authorship, and ownership.

Collaboration. Participants framed the relationship as hierarchical, positioning themselves as responsible for the overall creative vision while the AI carried out specific tasks. Several used managerial metaphors, describing themselves as a “CEO” (P4) or “creative director” (P9) directing an assistant who handles creative details

based on their instructions, sometimes reduced to a “word chooser” (P8). However, collaboration broke down when the system appeared to act independently of the creator’s intent. P10 described the process as frustrating when the AI ignored explicit instructions, saying that in those moments the system had “more of a say in the final outcome” than they did, while P7 remarked that simply clicking “generate” left them feeling “out of the loop.” As P3 summarized, the tool was “supposed to be a collaboration,” but at times felt like an argument with a collaborator who wanted to “do its own thing.”

Authorship. Substantial modifications often reduced artists’ sense of authorship and led to reluctance in delegating the writing process to AI. Because songwriting typically involves careful word choice, even small changes could make the result feel misaligned with the artist’s voice. For example, P2 explained that in songwriting “we choose every word very carefully,” so even changing “just one word feels like it’s a little off.” When the generated lyrics diverged further from the original song, participants sometimes felt disconnected from the outputs, or lost the sense of authorship. P10 remarked that the output was “very quickly becoming not my song, because it was using words that I didn’t even know.” This highlights the need to balance newly generated output with creator-specified constraints.

Ownership. Participants expressed various views about who owns AI-adapted lyrics.

P1 (AI engineer), who already uses AI in their creative workflows, maintained full ownership, viewing AI as a tool rather than a co-author.

Others described ownership as distributed across multiple actors shared between “me (the artist), the listener, and the AI” (P6). For instance, P4 suggested that ownership may lie with the listener, as their inputs shape the adaptation into an “audience-authored” version of the song. Similarly, participants described sharing ownership with the algorithm (P10), or the human developer (P7) and company (P8) who created the AI. Finally, some participants framed AI-generated adaptations as derivative works inspired by the creator’s style rather than authored by them. P9 compared the process to AI-generated visual art that is “Monet-inspired”—reflecting an artist’s influence without actually belonging to them—and argued that while the output might not fully belong to the songwriter, creators should still receive credit and compensation for work derived from their artistic style. Furthermore, incorporating audience information makes the adaptation reflect the listener’s persona, creating a sense that the song becomes “audience-authored” (P4), thereby introducing greater ambiguity in authorship and ownership.

8 Discussion

We discuss limitations of our approach, and outline directions for future research.

8.1 Content adaptation beyond lyrics

We instantiated our adaptive media authoring framework in the domain of songwriting. However, we anticipate advances in generative AI may enable other forms of adaptive multimedia. We hope to explore these richer forms of adaptive media and how the C³ framework can be instantiated in different domains.

Authoring adaptive songs. MultiVerse currently focuses on text (i. e., lyrics) and assumes a fixed melody with finalized lyrics. Future systems could allow creators to compose multiple melodic variations or entirely new tunes alongside lyrics. Supporting this would require tools for generating, adapting, and previewing musical content, while giving creators control over which melodic elements are preserved, modified, or recombined.

Authoring adaptive media beyond music. While we demonstrate our approach in songwriting, other media—such as film, audiobooks, and poetry—could also be adaptive. For example, comics could vary dialogue or visuals, and podcasts could adjust tone for different listeners. In such settings, creators face the same core challenge: specifying how content should change across contexts without directly reviewing each variation. This suggests that steering may be a useful abstraction for adaptive media more broadly, and C^3 offers a foundation for navigating this emerging design space.

8.2 Context acquisition and deployment

Adaptive media ultimately depends on both creator and audience adoption. While this paper focuses on how creators author adaptive experiences, an equally important direction for future work is understanding how audience context is acquired in deployed systems and how listeners respond to context-aware adaptations.

Envisioned deployment. Deployment requires deciding what audience context to use, whether that context should trigger personalization at all, and how personalized content should be presented to listeners. Audience context may be explicitly provided by listeners (e. g., an extended version of a user profile) or inferred from interaction history, sensors, or user models [67]. Context availability, however, does not imply that personalization should always occur. Adaptive systems may instead reserve personalization for moments when context meaningfully complements the creator’s intent or the listener’s situation, and otherwise present the original work.

Audience perspectives. Deployment decisions also raise questions of privacy and user agency. Future work may draw inspiration from work on privacy-aware adaptation techniques and privacy negotiation in XR to develop adaptive media systems that balance personalization with audience privacy preferences [61, 63]. More broadly, adaptive media should be evaluated not only by what context it can use, but by how listeners experience those choices: whether personalization feels meaningful, appropriate, intrusive, or authentically tied to the work. Future work should therefore examine which deployments listeners perceive as worthwhile, which contextual cues they find acceptable, and when the original work should remain unchanged.

8.3 Creator challenges for authoring adaptive media

Unlike static media, adaptive media is experienced through many runtime-generated variations. Because creators cannot directly review every possible adaptation, authoring becomes the task of specifying behavior over a potentially unbounded space of future

outputs. This introduces a fundamental challenge for adaptive media: how can creators meaningfully control outputs they will never directly observe?

From steering to explicit adaptation policies. One approach to addressing this challenge is to explore authoring explicit and verifiable adaptation policies. Most current AI-assisted creative tools rely on natural-language prompts to influence model behavior, providing limited guarantees that generated outputs will consistently satisfy creator intent. This becomes particularly problematic for adaptive media, where outputs are generated at runtime and cannot be exhaustively reviewed. MultiVerse explores one step in this direction by combining generative models with deterministic verification of creator-authored constraints. However, our study also revealed that not all constraints should be treated equally. For example, one songwriter observed that preserving a rhyme constraint forced the system toward lexically compatible words that conflicted with the intended emotional tone of the lyrics. This suggests that future authoring tools should allow creators to express different levels of commitment to adaptation policies, specifying which requirements must always hold, which are preferred but negotiable, and how conflicts among competing policies should be resolved. Supporting explicit prioritization and conflict management may become increasingly important as adaptive media grows in complexity and the space of possible adaptations expands.

Modeling the author. While richer adaptation policies can provide stronger guarantees over runtime behavior, manually specifying them may become increasingly burdensome as adaptive experiences grow in complexity. Rather than explicitly authoring every adaptation policy, creators may instead author a computational representation capable of making low-level creative decisions during runtime. During our study, some participants wanted to provide the system with much richer representations of their creative process—for example, conversations, drafts, or other materials that inspired a song—to better preserve artistic intent across future adaptations. One way to realize this vision is to treat authoring as a longitudinal process, allowing systems to gradually construct richer representations of artistic intent from interaction traces such as drafts, prior prompts, and refinement decisions (e.g., which variations are accepted, rejected, or edited) [29, 52, 68, 69]. Similar to recent work on conversational agents that act on a user’s behalf [10], such representations could enable creators to delegate creative decisions to the system while remaining aligned with their goals. Under this view, the published artifact is no longer only the work itself, but also the computational representation that continues shaping how the work evolves across audiences and contexts.

9 Conclusion

We presented C^3 , a creator-centered approach for steering adaptive media, and instantiated it in MultiVerse for adaptive lyric authoring. As dynamic media experiences expand beyond interactivity, new authoring paradigms will be needed to support creators in designing how their work responds to diverse consumption contexts. Through MultiVerse, we show that domain-specific tools can help creators steer adaptations toward outcomes that align with their artistic intent. More broadly, we hope our work encourages further

exploration of creator-centered tools for context-adaptive media across domains such as music, video, and other emerging forms of dynamic media.

References

- [1] Margareta Ackerman and David Loker. 2017. Algorithmic Songwriting with ALYSIA. In *Computational Intelligence in Music, Sound, Art and Design*, João Correia, Vic Ciesielski, and Antonios Liapis (Eds.). Springer International Publishing, Cham, 1–16.
- [2] Eytan Adar, Carolyn Gearig, Ayshwarya Balasubramanian, and Jessica Hullman. 2017. PersaLog: Personalization of News Article Content. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 3188–3200. doi:10.1145/3025453.3025631
- [3] Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A. Hearst, Andrew Head, and Kyle Lo. 2023. Paper Plain: Making Medical Research Papers Approachable to Healthcare Consumers with Natural Language Processing. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 74 (Sept. 2023), 38 pages. doi:10.1145/3589955
- [4] Stephen Brade, Bryan Wang, Mauricio Sousa, Gregory Lee Newsome, Sageev Oore, and Tovi Grossman. 2024. SynthScribe: Deep Multimodal Tools for Synthesizer Sound Retrieval and Exploration. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) (IUI '24). Association for Computing Machinery, New York, NY, USA, 51–65. doi:10.1145/3640543.3645158
- [5] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (UIST '23). Association for Computing Machinery, New York, NY, USA, Article 96, 14 pages. doi:10.1145/3586183.3606725
- [6] Antoine Caillon and Philippe Esling. 2021. RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. *arXiv preprint arXiv:2111.05011* (2021).
- [7] Alex Calderwood, John Joon Young Chung, Yuqian Sun, Melissa Roemmele, and Max Kreminski. 2025. Phraselette: A Poet's Procedural Palette. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 2701–2717. doi:10.1145/3715336.3735832
- [8] Chih-Ming Chen, Ming-Feng Tsai, Jen-Yu Liu, and Yi-Hsuan Yang. 2013. Using emotional context from article for contextual music recommendation. In *Proceedings of the 21st ACM International Conference on Multimedia* (Barcelona, Spain) (MM '13). Association for Computing Machinery, New York, NY, USA, 649–652. doi:10.1145/2502081.2502170
- [9] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. 2021. SemanticAdapt: Optimization-based Adaptation of Mixed Reality Layouts Leveraging Virtual-Physical Semantic Connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) (UIST '21). Association for Computing Machinery, New York, NY, USA, 282–297. doi:10.1145/3472749.3474750
- [10] Yi Fei Cheng, Hirokazu Shirado, and Shunichi Kasahara. 2025. Conversational Agents on Your Behalf: Opportunities and Challenges of Shared Autonomy in Voice Communication for Multitasking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [11] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (June 2014), 25 pages. doi:10.1145/2617588
- [12] Hyunsung Cho, Alexander Wang, Divya Kartik, Emily Liying Xie, Yukang Yan, and David Lindlbauer. 2024. Aupimize: Optimal Placement of Spatial Audio Cues for Extended 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 126, 14 pages. doi:10.1145/3654777.3676424
- [13] John Joon Young Chung and Eytan Adar. 2023. Artinter: AI-powered Boundary Objects for Commissioning Visual Arts. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) (DIS '23). Association for Computing Machinery, New York, NY, USA, 1997–2018. doi:10.1145/3563657.3595961
- [14] John Joon Young Chung and Max Kreminski. 2024. Patchview: LLM-powered Worldbuilding with Generative Dust and Magnet Visualization. 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 77, 19 pages. doi:10.1145/3654777.3676352
- [15] Anind K Dey. 2001. Understanding and using context. *Personal and ubiquitous computing* 5, 1 (2001), 4–7. doi:10.1007/s007790170019
- [16] Chris Donahue, Mina Lee, and Percy Liang. 2020. Enabling Language Models to Fill in the Blanks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 2492–2501. doi:10.18653/v1/2020.acl-main.225
- [17] Jesse Engel, Lamtharn (Hanoi) Hantrakul, Chenjie Gu, and Adam Roberts. 2020. DDSP: Differentiable Digital Signal Processing. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=B1x1ma4tDr>
- [18] João Marcelo Evangelista Belo, Mathias N. Lystbæk, Anna Maria Feit, Ken Pfeuffer, Peter Kán, Antti Oulasvirta, and Kaj Grönbæk. 2022. AUIT – the Adaptive User Interfaces Toolkit for Designing XR Applications. 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 48, 16 pages. doi:10.1145/3526113.3545651
- [19] Jonas Freiknecht and Wolfgang Effelsberg. 2020. Procedural Generation of Interactive Stories using Language Models (FDG '20). Association for Computing Machinery, New York, NY, USA, Article 97, 8 pages. doi:10.1145/3402942.3409599
- [20] Yue Fu, Michele Newman, Lewis Going, Qiuzi Feng, and Jin Ha Lee. 2025. Exploring the Collaborative Co-Creation Process with AI: A Case Study in Novice Music Production. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 1298–1312. doi:10.1145/3715336.3735829
- [21] Jacob Garbe, Max Kreminski, Ben Samuel, Noah Wardrip-Fruin, and Michael Mateas. 2019. StoryAssembler: an engine for generating dynamic choice-driven narratives. In *Proceedings of the 14th International Conference on the Foundations of Digital Games* (San Luis Obispo, California, USA) (FDG '19). Association for Computing Machinery, New York, NY, USA, Article 24, 10 pages. doi:10.1145/3337722.3337732
- [22] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing using Language Models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) (DIS '22). Association for Computing Machinery, New York, NY, USA, 1002–1019. doi:10.1145/3532106.3535333
- [23] Frederic Gmeiner, Nicolai Marquardt, Michael Bentley, Hugo Romat, Michel Pahud, David Brown, Asta Roseway, Nikolas Martelaro, Kenneth Holstein, Ken Hinckley, and Nathalie Riche. 2025. Intent Tagging: Exploring Micro-Prompting Interactions for Supporting Granular Human-GenAI Co-Creation Workflows. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 531, 31 pages. doi:10.1145/3706598.3713861
- [24] Junmin Gong, Sean Zhao, Sen Wang, Shengyuan Xu, and Joe Guo. 2025. Ace-step: A step towards music generation foundation model. *arXiv preprint arXiv:2506.00045* (2025).
- [25] Gemini Google. 2025. Nano Banana - Gemini AI Image Generator and Photo Editor. <https://gemini.google/overview/image-generation>. Accessed on November 5, 2025.
- [26] Daniel Green, Charlie Hargood, and Fred Charles. 2019. Define “authoring tool”: a survey of interactive narrative authoring tools. In *11th International Conference for Interactive Digital Storytelling: ICIDS 2018*.
- [27] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2017. Towards Pervasive Augmented Reality: Context-awareness in Augmented Reality. *IEEE transactions on visualization and computer graphics* 23, 6 (June 2017), 1706–1724. doi:10.1109/TVCG.2016.2543720
- [28] Aditya Gunturu, Yi Wen, Nandi Zhang, Jarin Thundathil, Rubaiat Habib Kazi, and Ryo Suzuki. 2024. Augmented Physics: Creating Interactive and Embedded Physics Simulations from Static Textbook Diagrams. 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 144, 12 pages. doi:10.1145/3654777.3676392
- [29] Noor Hammad. 2025. Towards Trail-Aware Digital Content Creation Tools. In *Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY Companion '25)*. Association for Computing Machinery, New York, NY, USA, 273–275. doi:10.1145/3744736.3749328
- [30] Noor Hammad, C. Ailie Fraser, Erik Harpstead, Jessica Hammer, and Mira Dontcheva. 2025. “It’s more of a vibe I’m going for”: Designing Text-to-Music Generation Interfaces for Video Creators. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 2738–2754. doi:10.1145/3715336.3735814
- [31] Violet Yinu Han, Hyunsung Cho, Kiyosu Maeda, Alexandra Ion, and David Lindlbauer. 2023. BlendMR: A Computational Method to Create Ambient Mixed Reality Interfaces. *Proc. ACM Hum.-Comput. Interact.* 7, ISS, Article 436 (Nov. 2023), 25 pages. doi:10.1145/3626472
- [32] Melanie Heck, Janick Edinger, Jonathan Bünemann, and Christian Becker. 2021. The Subconscious Director: Dynamically Personalizing Videos Using Gaze Data. In *Proceedings of the 26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) (IUI '21). Association for Computing Machinery, New York, NY, USA, 98–108. doi:10.1145/3397481.3450679
- [33] Daniel Howe. 2020. RiTa: AI-free tools for natural language and generative writing. <https://rednoise.org/rita/>. Accessed: 2026-03-14.
- [34] Cheng-Zhi Anna Huang, Tim Coolijmans, Adam Roberts, Aaron Courville, and Douglas Eck. 2019. Counterpoint by convolution. *arXiv preprint arXiv:1903.07227*

- (2019).
- [35] Jeremy Zhengqi Huang, Caluã de Lacerda Patata, Liang-Yuan Wu, and Dhruv Jain. 2025. CapTune: Adapting Non-Speech Captions With Anchored Generative Models. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25)*. Association for Computing Machinery, New York, NY, USA, Article 74, 18 pages. doi:10.1145/3663547.3746346
 - [36] Mina Huh, Ding Li, Kim Pimmel, Hujung Valentina Shin, Amy Pavel, and Mira Dontcheva. 2025. VideoDiff: Human-AI Video Co-Creation with Alternatives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1143, 19 pages. doi:10.1145/3706598.3713417
 - [37] Nicholas Jennings, Han Wang, Isabel Li, James Smith, and Bjoern Hartmann. 2024. What's the Game, then? Opportunities and Challenges for Runtime Behavior Generation. 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 106, 13 pages. doi:10.1145/3654777.3676358
 - [38] Brett R. Jones, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2013. IllumiRoom: peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Paris, France) (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 869–878. doi:10.1145/2470654.2466112
 - [39] Mohamed Kari, Tobias Grosse-Puppenthal, Alexander Jagaciak, David Bethge, Reinhard Schütte, and Christian Holz. 2021. SoundsRide: Affordance-Synchronized Music Mixing for In-Car Audio Augmented Reality. In *The 34th Annual ACM Symposium on User Interface Software and Technology (Virtual Event, USA) (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 118–133. doi:10.1145/3472749.3474739
 - [40] Mohamed Kari, Reinhard Schütte, and Raj Sodhi. 2023. Scene Responsiveness for Visuotactile Illusions in Mixed Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (San Francisco, CA, USA) (UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 84, 15 pages. doi:10.1145/3586183.3606825
 - [41] Taewook Kim, Dhruv Agarwal, Jordan Ackerman, and Manaswi Saha. 2025. Steering AI-driven Personalization of Scientific Text for General Audiences. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article CSCW479 (Oct. 2025), 28 pages. doi:10.1145/3757660
 - [42] Taewook Kim, Hyomin Han, Eytan Adar, Matthew Kay, and John Joon Young Chung. 2024. Authors' Values and Attitudes Towards AI-bridged Scalable Personalization of Creative Language Arts. 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 31, 16 pages. doi:10.1145/3613904.3642529
 - [43] Taewook Kim, Jung Soo Lee, Zhenhui Peng, and Xiaojuan Ma. 2019. Love in Lyrics: An Exploration of Supporting Textual Manifestation of Affection in Social Messaging. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 79 (Nov. 2019), 27 pages. doi:10.1145/3359181
 - [44] Yewon Kim, Sung-Ju Lee, and Chris Donahue. 2025. Amuse: Human-AI Collaborative Songwriting with Multimodal Inspirations. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 187, 28 pages. doi:10.1145/3706598.3713818
 - [45] Tomas Lawton, Francisco J Ibarrola, Dan Ventura, and Kazjon Grace. 2023. Drawing with Reframer: Emergence and Control in Co-Creative AI. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (Sydney, NSW, Australia) (IUI '23)*. Association for Computing Machinery, New York, NY, USA, 264–277. doi:10.1145/3581641.3584095
 - [46] Mackenzie Leake, Abe Davis, Anh Truong, and Maneesh Agrawala. 2017. Computational video editing for dialogue-driven scenes. *ACM Trans. Graph.* 36, 4, Article 130 (July 2017), 14 pages. doi:10.1145/3072959.3073653
 - [47] Michael Lebowitz. 1985. Story-telling as planning and learning. *Poetics* 14, 6 (1985), 483–502.
 - [48] Hsin-Pei Lee, Jhih-Sheng Fang, and Wei-Yun Ma. 2019. iComposer: An Automatic Songwriting System for Chinese Popular Music. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, Waleed Ammar, Annie Louis, and Nasrin Mostafazadeh (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 84–88. doi:10.18653/v1/N19-4015
 - [49] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A. Alghamdi, Tal August, Avinash Bhat, Madiha Zahrah Choksi, Senjuti Dutta, Jin L.C. Guo, Md Naimul Hoque, Yewon Kim, Simon Knight, Seyed Parsa Neshaei, Antonette Shibani, Disha Shrivastava, Lila Shroff, Agnia Sergeyuk, Jessi Stark, Sarah Stermann, Sitong Wang, Antoine Bosselut, Daniel Buschek, Joseph Chee Chang, Sherol Chen, Max Kreminski, Joonsuk Park, Roy Pea, Eugenia Ha Rim Rho, Zejiang Shen, and Pao Siangliulue. 2024. 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) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1054, 35 pages. doi:10.1145/3613904.3642697
 - [50] H Lieberman and T Selker. 2000. Out of context: Computer systems that adapt to, and learn from, context. *IBM systems journal* 39, 3.4 (2000), 617–632. doi:10.1147/sj.393.0617
 - [51] David Chuan-En Lin, Hyeonsu B. Kang, Nikolas Martelaro, Aniket Kittur, Yan-Ying Chen, and Matthew K. Hong. 2025. Inkspire: Supporting Design Exploration with Generative AI through Analogical Sketching. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 427. doi:10.1145/3706598.3713397
 - [52] David Chuan-En Lin and Nikolas Martelaro. 2021. Learning personal style from few examples. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. 1566–1578.
 - [53] Xingyu Bruce Liu, Mira Dontcheva, and Dingzeyu Li. 2026. A Text-Native Interface for Generative Video Authoring. *arXiv preprint arXiv:2603.09072* (2026).
 - [54] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J. Cai. 2020. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376739
 - [55] Ryan Louie, Jesse Engel, and Cheng-Zhi Anna Huang. 2022. Expressive Communication: Evaluating Developments in Generative Models and Steering Interfaces for Music Creation. In *Proceedings of the 27th International Conference on Intelligent User Interfaces (Helsinki, Finland) (IUI '22)*. Association for Computing Machinery, New York, NY, USA, 405–417. doi:10.1145/3490099.3511159
 - [56] Zhuoran Lu, Qian Zhou, and Yi Wang. 2025. WhatELSE: Shaping Narrative Spaces at Configurable Level of Abstraction for AI-bridged Interactive Storytelling. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 333, 18 pages. doi:10.1145/3706598.3713363
 - [57] Rudolf Mayer and Andreas Rauber. 2011. Musical genre classification by ensembles of audio and lyrics features. In *Proceedings of international conference on music information retrieval*. 675–680.
 - [58] William Moylan. 2020. *Recording analysis: How the record shapes the song*. Focal Press.
 - [59] Michele Newman, Lidia Morris, and Jin Ha Lee. 2023. Human-AI Music Creation: Understanding the Perceptions and Experiences of Music Creators for Ethical and Productive Collaboration.. In *ISMIR*. 80–88.
 - [60] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (San Francisco, CA, USA) (UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 2, 22 pages. doi:10.1145/3586183.3606763
 - [61] Shwetha Rajaram, Jiasi Chen, and Michael Nebeling. 2025. Privacy Equilibrium: Balancing Privacy Needs in Dynamic Multi-User Augmented Reality Scenarios. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, Article 151, 24 pages. doi:10.1145/3746059.3747783
 - [62] Shwetha Rajaram, Nels Numan, Balasaravanan Thoravi Kumaravel, Nicolai Marquardt, and Andrew D Wilson. 2024. BlendScape: Enabling End-User Customization of Video-Conferencing Environments through Generative AI. 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 40, 19 pages. doi:10.1145/3654777.3676326
 - [63] Shwetha Rajaram, Macarena Peralta, Janet G. Johnson, and Michael Nebeling. 2025. Exploring the Design Space of Privacy-Driven Adaptation Techniques for Future Augmented Reality Interfaces. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1226, 19 pages. doi:10.1145/3706598.3713320
 - [64] M.O. Riedl and R.M. Young. 2006. From linear story generation to branching story graphs. *IEEE Computer Graphics and Applications* 26, 3 (2006), 23–31. doi:10.1109/MCG.2006.56
 - [65] Mark O Riedl and R Michael Young. 2006. Story planning as exploratory creativity: Techniques for expanding the narrative search space. *New Generation Computing* 24, 3 (2006), 303–323.
 - [66] Markus Schedl and Dominik Schnitzer. 2014. Location-aware music artist recommendation. In *International conference on multimedia modeling*. Springer, 205–213.
 - [67] Omar Shaikh, Shardul Sapkota, Shan Rizvi, Eric Horvitz, Joon Sung Park, Diyi Yang, and Michael S. Bernstein. 2025. Creating General User Models from Computer Use. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, Article 35, 23 pages. doi:10.1145/3746059.3747722
 - [68] Amy Smith, Barrett R Anderson, Jasmine Tan Otto, Isaac Karth, Yuqian Sun, John Joon Young Chung, Melissa Roemmele, and Max Kreminski. 2025. Scaling Analysis of Creative Activity Traces via Fuzzy Linkography. In *Proceedings of the*

- Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 469, 10 pages. doi:10.1145/3706599.3720184
- [69] Amy Smith, Barrett R Anderson, Jasmine Tan Otto, Isaac Karth, Yuqian Sun, John Joon Young Chung, Melissa Roemmele, and Max Kreminski. 2025. Fuzzy Linkography: Automatic Graphical Summarization of Creative Activity Traces. In *Proceedings of the 2025 Conference on Creativity and Cognition (C&C '25)*. Association for Computing Machinery, New York, NY, USA, 637–650. doi:10.1145/3698061.3726915
- [70] Yusen Sun, Liangyou Li, Qun Liu, and Dit-Yan Yeung. 2023. SongRewriter: A Chinese Song Rewriting System with Controllable Content and Rhyme Scheme. In *Findings of the Association for Computational Linguistics: ACL 2023*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 12863–12880. doi:10.18653/v1/2023.findings-acl.814
- [71] Yuying Tang, Xinyi Chen, Haotian Li, Xing Xie, Xiaojuan Ma, and Huamin Qu. 2026. DuoDrama: Supporting Screenplay Refinement Through LLM-Assisted Human Reflection. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, Article 1078, 20 pages. doi:10.1145/3772318.3790568
- [72] Yujie Tao and Pedro Lopes. 2022. Integrating Real-World Distractions into Virtual Reality. 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 5, 16 pages. doi:10.1145/3526113.3545682
- [73] John Thickstun, David Leo Wright Hall, Chris Donahue, and Percy Liang. 2024. Anticipatory Music Transformer. *Transactions on Machine Learning Research* (2024).
- [74] Fang-Duo Tsai, Shih-Lun Wu, Haven Kim, Bo-Yu Chen, Hao-Chung Cheng, and Yi-Hsuan Yang. 2024. Audio prompt adapter: Unleashing music editing abilities for text-to-music with lightweight finetuning. *arXiv preprint arXiv:2407.16564* (2024).
- [75] Alexander Wang, Yi Fei Cheng, and David Lindlbauer. 2024. MARingBA: Music-Adaptive Ringtones for Blended Audio Notification Delivery. 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 729, 15 pages. doi:10.1145/3613904.3642376
- [76] Alexander Wang, Chris Donahue, and Dhruv Jain. 2025. RISE: Music Rearrangement for Realtime Intensity Synchronization With Exercise. In *Proceedings of the 26th International Society for Music Information Retrieval Conference*. ISMIR, 34–41. doi:10.5281/zenodo.17706321
- [77] Alexander Wang, David Lindlbauer, and Chris Donahue. 2024. Towards Music-Aware Virtual Assistants. 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 127, 14 pages. doi:10.1145/3654777.3676416
- [78] Bryan Wang, Yuliang Li, Zhaoyang Lv, Haijun Xia, Yan Xu, and Raj Sodhi. 2024. LAVE: LLM-Powered Agent Assistance and Language Augmentation for Video Editing. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) (IUI '24). Association for Computing Machinery, New York, NY, USA, 699–714. doi:10.1145/3640543.3645143
- [79] Sitong Wang, Anh Truong, Lydia B Chilton, and Dingzeyu Li. 2026. Rewriting Video: Text-Driven Reauthoring of Video Footage. *arXiv preprint arXiv:2601.08565* (2026).
- [80] Yi Wang, Qian Zhou, and David Ledo. 2024. StoryVerse: Towards Co-authoring Dynamic Plot with LLM-based Character Simulation via Narrative Planning. In *Proceedings of the 19th International Conference on the Foundations of Digital Games* (Worcester, MA, USA) (FDG '24). Association for Computing Machinery, New York, NY, USA, Article 55, 4 pages. doi:10.1145/3649921.3656987
- [81] Kento Watanabe, Yuichiroh Matsubayashi, Satoru Fukayama, Masataka Goto, Kentaro Inui, and Tomoyasu Nakano. 2018. A Melody-Conditioned Lyrics Language Model. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Marilyn Walker, Heng Ji, and Amanda Stent (Eds.). Association for Computational Linguistics, New Orleans, Louisiana, 163–172. doi:10.18653/v1/N18-1015
- [82] Kento Watanabe, Yuichiroh Matsubayashi, Kentaro Inui, Tomoyasu Nakano, Satoru Fukayama, and Masataka Goto. 2017. LyriSys: An Interactive Support System for Writing Lyrics Based on Topic Transition. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) (IUI '17). Association for Computing Machinery, New York, NY, USA, 559–563. doi:10.1145/3025171.3025194
- [83] Shih-Lun Wu, Chris Donahue, Shinji Watanabe, and Nicholas J. Bryan. 2024. Music ControlNet: Multiple Time-Varying Controls for Music Generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 32 (2024), 2692–2703. doi:10.1109/TASLP.2024.3399026
- [84] Yusong Wu, Ke Chen, Tianyu Zhang, Yuchen Hui, Taylor Berg-Kirkpatrick, and Shlomo Dubnov. 2023. Large-Scale Contrastive Language-Audio Pretraining with Feature Fusion and Keyword-to-Caption Augmentation. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 1–5. doi:10.1109/ICASSP49357.2023.10095969
- [85] Haijun Xia. 2020. Crosspower: Bridging Graphics and Linguistics. 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, 722–734. doi:10.1145/3379337.3415845
- [86] Lilian Yang. 2022. *How K-Pop Apps Create the Illusion of Private Messaging with Celebrities*. Nielsen Norman Group. Retrieved July 9, 2026 from <https://www.nngroup.com/articles/kpop-private-messaging/>
- [87] Eva Zangerle, Martin Pichl, and Markus Schedl. 2018. Culture-Aware Music Recommendation. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization* (Singapore, Singapore) (UMAP '18). Association for Computing Machinery, New York, NY, USA, 357–358. doi:10.1145/3209219.3209258
- [88] Eva Zangerle, Martin Pichl, and Markus Schedl. 2020. User models for culture-aware music recommendation: fusing acoustic and cultural cues. *Transactions of the International Society for Music Information Retrieval* 3, 1 (2020).

Table 1: Participant musical backgrounds. FT = years of formal musical training; SW = years of songwriting experience.

ID	FT	SW	Musical Practice	Genres
P1	10	5	EDM producer; pianist, drummer, and DJ; experience with AI-assisted songwriting	Progressive/Future House, EDM, Pop/Punk
P2	16	10	Graduate screen scoring student; composer for indie games and films; released album; active songwriter across genres	Electronic Pop, Alt Pop, Indie
P3	1	4	Vocal training; hobby songwriter and producer	Pop, R&B
P4	14	5	Violinist with orchestral and jazz background; writes instrumentals and lyric adaptations	Pop
P5	0	5	Collegiate a cappella performer; co-wrote and produced albums for collaborators	Bedroom Jazz Fusion
P6	4	16	Primary songwriter and lyricist for two bands; producer, engineer, and performer	Indie Rock, Synth-pop, Electronic, Experimental
P7	10	15	Songwriting and production for multimedia projects (theater, animation)	Indie
P8	6	5	Live performer; released original songs; narrative-focused lyric songwriter	Folk Rock
P9	16	5	Electronic music major; classical pianist; commissioned composer working professionally in music and film	Pop, R&B, Alternative, Cinematic
P10	15	12	Audio engineering and music production professor at university; active songwriter (50–100 songs); multi-instrumentalist and music educator	Pop, Rock, Indie

ORIGINAL LYRICS

[Verse 1]
 We've known each other for so long
 Your heart's been aching, but you're too shy to say it
 Inside, we both know what's been going on
 We know the game, and we're gonna play it

AUTHOR ADAPTATION INSTRUCTIONS [CONTROL DOCUMENT]

NOTE:
 This instruction document is the actual file used to control dynamic output.

INSTRUCTION FOR AI:
 "Rewrite the lyrics based on its intended audience [audience information] and the [control document] while adhering the following instructions:"

keep the number of syllables consistent.
 The line ending in "long" should rhyme with the one ending with "on"

[AUDIENCE INFORMATION]

CONTEXT:
 Select a persona from Group 1 to define the audience context.

Elena, 30F, Game Designer, Berlin

NARRATIVE PROFILE:
 Elena, 30, is an indie game designer in Berlin who lives between glitch aesthetics and deep-flow coding states, often finding her most vivid social interactions through late-night Discord channels.

ID REF:
 persona_01

Generate Adaptation

LYRIC PREVIEW
 Refined for: Elena

Our builds have buffered for so long
 Your code is crashing, but you're too shy to say it
 Internal pings show what is going on
 We know the game, and we're gonna play it

Download Result (JSON)

Figure 12: Interface for prompting-based workflow.

Question	Mean MultiVerse	Mean Prompt
I would be happy to use this system or tool on a regular basis.	4.5 (SD = 1.3)	4.1 (SD = 2.1)
I enjoyed using the system or tool.	5.0 (SD = 1.3)	5.0 (SD = 1.6)
It was easy for me to explore many different ideas, options, designs, or outcomes using this system or tool.	5.2 (SD = 1.6)	4.6 (SD = 1.8)
The system or tool was helpful in allowing me to track different ideas, outcomes, or possibilities.	4.8 (SD = 1.5)	3.6 (SD = 1.8)
I was able to be very creative while doing the activity inside this system or tool.	4.5 (SD = 1.6)	4.8 (SD = 1.7)
The system or tool allowed me to be very expressive.	4.2 (SD = 1.8)	3.9 (SD = 1.6)
My attention was fully tuned to the activity, and I forgot about the system or tool that I was using.	4.1 (SD = 1.9)	3.9 (SD = 2.1)
I became so absorbed in the activity that I forgot about the system or tool that I was using.	3.5 (SD = 1.8)	3.4 (SD = 1.8)
I was satisfied with what I got out of the system or tool.	4.0 (SD = 1.8)	4.4 (SD = 1.6)
What I was able to produce was worth the effort I had to exert to produce it.	4.1 (SD = 1.7)	4.5 (SD = 1.4)
I was able to effectively communicate what I wanted to the system.	4.3 (SD = 1.3)	4.3 (SD = 1.7)
I was able to steer the system towards output that was aligned with my goals.	4.7 (SD = 2.1)	5.0 (SD = 1.7)
At times, I felt that the system was steering me towards its own goals.	4.5 (SD = 1.7)	4.6 (SD = 1.7)
I felt the lyrics created were mine.	2.9 (SD = 1.8)	2.8 (SD = 1.3)
The music created was totally due to my contributions (as opposed to the system's contributions).	3.3 (SD = 1.5)	3.8 (SD = 1.4)
I felt like I was collaborating with the system.	4.7 (SD = 2.1)	5.1 (SD = 1.7)
I believe the generated lyrics fits well with the original melody.	4.2 (SD = 1.8)	4.5 (SD = 2.0)
I believe the generated lyrics preserved the original intent of the song.	4.0 (SD = 1.7)	3.8 (SD = 1.8)
I believe the generated lyrics adapted appropriately for different audiences.	4.9 (SD = 1.2)	4.5 (SD = 1.6)
I believe the generated lyrics are of high quality.	3.3 (SD = 1.6)	4.0 (SD = 1.3)
I am satisfied with the generated lyrics overall.	3.4 (SD = 1.9)	3.8 (SD = 1.9)
I felt that the system helped me make informed decisions during the authoring process.	4.8 (SD = 1.4)	3.6 (SD = 1.9)
I could reasonably predict how the system would adapt the content after I finished authoring it.	4.8 (SD = 1.5)	3.8 (SD = 1.8)
I felt I had meaningful control over how the lyrics would adapt to different audiences.	4.9 (SD = 1.6)	3.9 (SD = 1.6)
I clearly understood how my controls affected the system's adaptive behavior.	5.0 (SD = 1.2)	4.0 (SD = 1.8)
I trust the system to generate acceptable versions of the content that I might never directly review.	3.0 (SD = 1.5)	2.7 (SD = 1.6)
I believe the system can effectively generate adaptations for a large number of different contexts.	4.5 (SD = 1.4)	3.5 (SD = 2.1)

Table 2: Survey questions and mean ratings for MultiVerse and prompting.

Table 3: Baseline prompt template.

Prompt = You are a songwriter, take the original lyrics and vary the lyrics while adhering to the instructions below:
 Rewrite the lyrics based on its intended audience \${audienceInfo} || '[audience information]'} and the [control document]
 while adhering the following instructions: \${instructions}
 Original Lyrics: \${lyrics}
 only output the lyrics without any additional text.