The Ghost in the Keys: A Disklavier Demo for Human-AI Musical Co-Creativity

Louis Bradshaw[1*] **Alexander Spangher**[2*] **Stella Biderman**[3] **Simon Colton**[1]

[1]Queen Mary University of London
[2]Stanford University
[3]EleutherAI
louis.bradshaw@qmul.ac.uk

Abstract

While generative models for music composition are increasingly capable, their adoption by musicians is hindered by text-prompting, an asynchronous workflow disconnected from the embodied, responsive nature of instrumental performance. To address this, we introduce *Aria-Duet*[1], an interactive system facilitating a real-time musical duet between a human pianist and *Aria*, a state-of-the-art generative model, using a Yamaha Disklavier as a shared physical interface. The framework enables a turn-taking collaboration: the user performs, signals a handover, and the model generates a coherent continuation performed acoustically on the piano. Beyond describing the technical architecture enabling this low-latency interaction, we analyze the system's output from a musicological perspective, finding the model can maintain stylistic semantics and develop coherent phrasal ideas, demonstrating that such embodied systems can engage in musically sophisticated dialogue and open a promising new path for human-AI co-creation.

1 Introduction

The adoption of modern AI music tools suggests a notable trend. While tools for passive tasks like source separation appear to be widely adopted into the creative workflows of producers, models that generate core compositional content have seemingly seen slower uptake among classically trained musicians and composers. This perceived gap is not, we argue, because musicians are averse to ceding creative control; it is caused, rather, by the *mode of interaction* presented by current AI models for musical composition. The paradigm of asynchronous text-prompting and speed-bottlenecked iteration is fundamentally at odds with the embodied, responsive, and often non-verbal feedback loops that define social *musicking* [Small, 1998] and *creative flow* [Csikszentmihalyi, 1990].

Aiming to bridge this gap for pianist-composers, we introduce *Aria-Duet*, an interactive system for real-time musical duets between a pianist and generative model. The system's physical interface is a Yamaha Disklavier, an acoustic piano capable of both capturing and physically playing performances through a MIDI connection. A musician plays on the instrument while the model listens; then, upon a *takeover signal*, the system responds by generating and playing a continuation on the same keys in real-time. This interaction is powered by *Aria* [Bradshaw et al., 2025], an autoregressive transformer [Vaswani et al., 2017, Huang et al., 2018] trained to compose expressive piano continuations one note at a time.

Aria-Duet is designed to recenter the artist's creative agency by restoring familiar feedback loops and enabling *experimental play*. This interaction is facilitated by the model's ability to adapt to a broad

[*]Equal contribution
[1]Available at: https://github.com/eleutherai/aria

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Creative AI Track.

range of musical styles, vocabularies, and forms. However, realizing an experience that feels genuinely fluid and engaging is not merely a matter of connecting a model to an instrument. As we will detail, the success of this interaction hinges on addressing key design challenges, including minimizing takeover latency, ensuring musical coherence, and maintaining accurate acoustic playback.

We conclude by demonstrating and providing a musicological analysis of the system's output. By examining our system's responses to varied prompts, we highlight its ability to maintain stylistic semantics, develop coherent phrasal ideas, and even engage in multi-voice dialogue. Overall, our explorations contribute toward a practical blueprint for designing real-time, interactive AI systems that augment the human creative process.

2 Related Work

Our work builds on decades of research into human-computer musical co-creation. Early interactive systems, such as George Lewis's *Voyager* [Lewis, 1999] and Biles's *GenJam* [Biles, 1994] pioneered real-time interaction using rule-based grammars or genetic algorithms. While inventive, these systems were often limited to pre-programmed styles and struggled to adapt to unexpected musical inputs from the performer.

The subsequent wave of *statistical style-learning* companions enabled systems to learn directly from a performer's input. The most influential of these, Pachet's *Continuator* [Pachet, 2003], introduced a back-and-forth keyboard partnership using Markov models that directly inspires our system's interactive paradigm. However, like other systems of its era (e.g., *BoB* [Thom, 2000], *OMax* [Assayag et al., 2006]), its reliance on methods with short context windows resulted in compositions which often failed to capture the deeper phrasal and structural logic typical of human-composed music.

The deep learning era brought architectures capable of modeling long-range dependencies. Systems like Google's *AI Duet* (using LSTMs [Eck, 2017]) and later Transformer-based models like *Music Transformer* [Huang et al., 2018] and *MuseNet* [OpenAI, 2019] demonstrated the ability to generate multi-bar structures and richer harmonic vocabularies. Recent work on interactive systems, such as *ReaLJam* [Scarlatos et al., 2025] and *Jam Bot* [Blanchard et al., 2025], have explored applications of modern neural models in real-time contexts. However, a gap remains: systems often sidestep the critical engineering challenges that are essential for a truly fluid, embodied interaction through acoustic instruments (e.g., Disklavier pianos). *Aria-Duet* directly addresses this gap by integrating a state-of-the-art model with a system meticulously designed for low-latency, acoustic performance.

3 System Design

Aria-Duet is comprised of two primary components: (1) a generative model for piano performance, adapted from *Aria* [Bradshaw et al., 2025], used to generate creative and expressive continuations of a user's performance on the Disklavier, and (2) a real-time engine that manages the user control flow, input/output, and real-time inference. In this section, we outline the design and implementation of each component.

3.1 Generative Model

Our system's continuation is generated by a model finetuned from *Aria* [Bradshaw et al., 2025], an autoregressive transformer model designed to model expressive symbolic piano performances (i.e., on the note-level). *Aria* is particularly well-suited for this application due to its training and tokenization scheme. It was pretrained on a refined subset of *Aria-MIDI* [Bradshaw and Colton, 2025], a large-scale (100k+ hours) dataset of solo piano music spanning a wide range of genres and styles. This dataset was curated using a transcription model that was itself trained on paired audio and MIDI recordings from a Disklavier [Hawthorne et al., 2018]. This creates a direct correspondence between the model's training data and the Disklavier-based I/O of our system. *Aria* employs a note-centric tokenizer that quantizes musical events with a fine-grained resolution, generating continuations by performing next-token (i.e., next-note) prediction in an iterative process.

In a preliminary version of our system, we used the native pretrained *Aria* model for generation. However, informal testing with pianists revealed two shortcomings that detracted from the co-creative experience. First, the model lacked explicit sustain pedal tokens, instead simulating sustain with

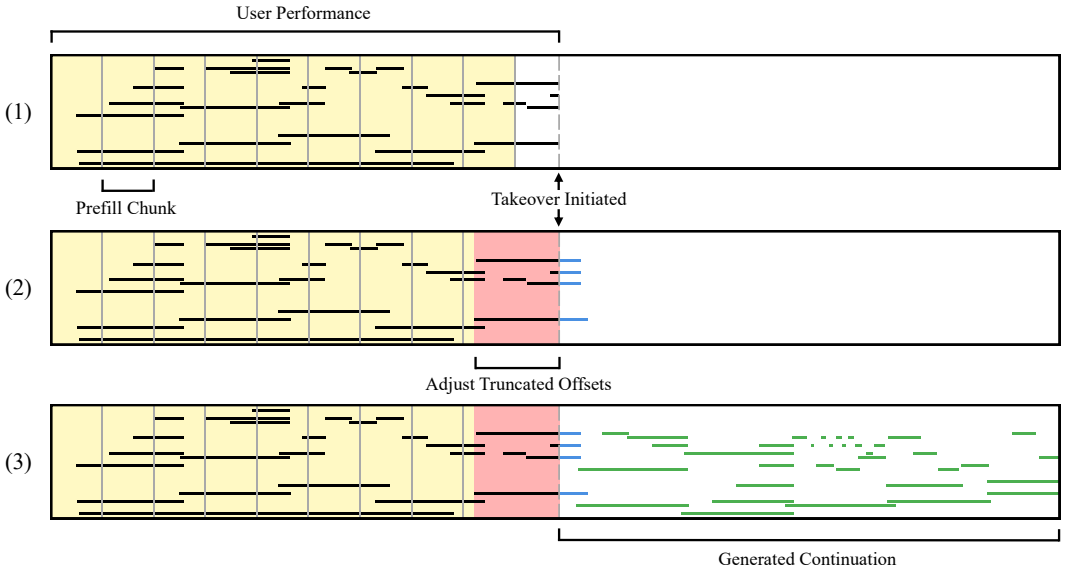

Figure 1: An illustration of the KV-Cache management for real-time, low-latency operation. (1) *Listen*: As the user plays, the received context is proactively and continuously prefilled into the model's KV-Cache in chunks. (2) *Takeover*: Upon a takeover signal, the system finalizes the input, prefilling any missing context and speculatively re-evaluating the durations of any hanging notes (seen in blue), ensuring a seamless transition and preparing the KV-Cache. (3) *Generate*: The model then begins generating a musical continuation note-by-note, streaming the result to the Disklavier.

immediate note retriggerings. While functionally equivalent in software, pianists universally found this approach jarring when played back on the Disklavier, which requires a gap to retrigger notes. One pianist also noted that this simulation only sustained notes, missing the acoustic resonance created by lifting the dampers. Second, due to the overrepresentation of popular works in the original training data, the model would often complete a famous theme when prompted. Classically trained pianists, in particular, found this frustrating, as it disrupted their natural tendency to use well-known repertoire as a starting point. To address these issues, we post-trained the model on a high-quality deduplicated subset of the *Aria-MIDI* dataset that included explicit pedal-on and pedal-off tokens. This single intervention addressed both problems, enabling the system to control the pedal while mitigating its tendency toward compositional memorization.

3.2 Real-time Engine

Our system's design is built upon an embodied, turn-based interaction model facilitated entirely by a Disklavier piano. The operational flow is as follows: a pianist begins by playing, and their performance is captured and routed to a computer as a stream of MIDI. To cede control to the generative model, the pianist presses the left pedal (*una corda*), which serves as the takeover signal. This triggers the system, running on a connected Apple Silicon device, to pre-process the performance and prompt the model to generate a musical continuation, which is streamed back to the Disklavier and performed on the keys in real-time. The pianist can reclaim control at any point by re-pressing the pedal, enabling an alternating musical dialogue that preserves the musical context. This control scheme aims to be intuitive, mapping system-level commands to familiar physical actions; however, to keep the interaction frictionless, several key engineering challenges must be addressed.

Response Latency. The inference process for autoregressive transformers involves two distinct computational phases: *prefill* and *decoding*. During prefill, the model processes the entire input prompt in parallel, a compute-bound (FLOPs) operation that populates a key-value (KV) cache with the context's attention states. Subsequently, the decoding phase generates the output one token at a time in an iterative process that is memory-bandwidth-bound. This presents a key challenge for real-time interaction on consumer hardware like Apple Silicon. While this hardware's high memory bandwidth is well-suited for fast decoding, the compute-bound prefill phase creates a significant

bottleneck, introducing an unacceptable lag of 1000-2000ms between the user's takeover signal and the model's first note. To mitigate this latency, we implement a *continuous prefill* strategy that proactively updates the KV-cache in small chunks as the user plays. This distributes the computational load over time, virtually eliminating the prefill-induced delay at the moment of transition.

Transition Coherence. Beyond minimizing the *time-to-first-note*, the musical coherence of the transition is equally important to the user experience. The core challenge stems from the fact that the model, trained to continue any input, is highly sensitive to the musical context immediately preceding the takeover. A common scenario where this manifests is when the user initiates a transition while notes are still held or sustained by the pedal. Due to the tokenizer's design, the model must be provided with complete note information, including durations, before it can predict new notes. A naive approach would be to force-end all active notes at the transition point. This, however, provides a 'truncated' context that often corrupts the model's predictions, causing it to generate similarly abrupt, staccato phrases. To remedy this, our system speculatively reevaluates the durations of notes truncated by the transition, filling these corrected durations into the model's KV-cache before generating a continuation. While this process adds a small latency overhead of ~100-200ms, it is vital for ensuring a smooth and musically coherent transition. Figure 1 illustrates the state of the KV-cache and how it is modified during these phases of operation.

Disklavier Playback. Translating the model's generated output into an accurate performance requires addressing the physical limitations of the Disklavier. Unlike for software synthesizers, the electromechanical action of a piano introduces two challenges: velocity-dependent note-on latency, where louder notes sound with different delays than softer ones; and mechanical conflicts, where a key cannot be retriggered before its action has physically reset. The Disklavier's native *playback mode* resolves these issues by using a buffer, but at the cost of a fixed 500ms latency that detracts from interactive use. Our system circumvents this by implementing a custom, zero-latency streaming layer that modifies the playback schedule in real-time. Instead of buffering, it makes two just-in-time adjustments: First, it schedules the send-time for each note-on message to account for manually-calibrated velocity-specific latency, and only articulates notes whose scheduled time arrives before exceeding a staleness threshold. Second, to prevent re-articulation errors, the system retrospectively modifies the send-time of a pending note-off message if a new note-on for the same pitch is generated before the off-message has been sent. This dynamic rescheduling enforces the necessary physical gap between notes by altering the timing of a future event, rather than by introducing a processing delay.

4 Demonstration

(a) Musician prompts *Aria-Duet*.

(b) *Aria-Duet* produces a response.

Figure 2: Musician interacts with the *Aria-Duet* system using a Disklavier.

In this section, we analyze the musicality of Aria-Duet's generated performances. Using a three-minute video demonstration[2], we explore how the system responds to a variety of prompts from a human pianist. In the demonstration are prompts to the following songs: *Carmen Overture* (Georges Bizet, 1875), *Ave Maria* (Franz Schubert, 1825; arranged by Franz Liszt, 1846), *Amazing Grace* (William Walker, 1835), *The Entertainer* (Scott Joplin, 1902), *Freygish Fino* (Noah Smith, 2024), *Spanish Song* (original composition, 2025), *Piano Concerto No. 2* (Sergei Rachmaninoff, 1901), and *Hungarian Rhapsody No. 6* (Franz Liszt, 1847).

***Semantics*: Vocabularies and Prosody.** According to Langer [1942], *vocabularies* emerge across modalities to meet the expressive needs of social cultures. *Musical* vocabularies emerge *both within songs and across genres* as *small motifs and patterns emerge and vary* [Schoenberg and Stein, 1984, Narmour, 1992]. Maintaining a piece's musical vocabulary is crucial for maintaining its coherence for the listener, thus, we explore how well *Aria* can maintain distributional semantics in given prompts.

[2]Available at: https://youtu.be/8s3V922h3CU

In our demonstration, we examine how *Aria* responds to prompts using less common tonalities, rhythms and chordal progressions[3]: for *Aria* to demonstrate a deep understanding of vocabulary, it must generate continuations adhering to these semantics and not simply revert to modal norms.

Spanish Song and *Freygish Fino* make use of the Freygish (Phrygian dominant) and Phrygian scales, which are relatively uncommon in mainstream Western tonal music compared to the diatonic major or minor scales. As can be seen in both instances, Aria is naturally able to identify and improvise within the appropriate scales; in *Freygish Fino*, it even continues and expands on the *Andalusian cadence*, a traditional Klezmer chord progression. *The Entertainer* offers an additional example of semantic consistency: *The Entertainer* uses the ragtime rhythmic structure, which consists of syncopation patterns and emphases on the 2/4 beats – both of which are outside standard Western classical music. *Aria* continues to maintain this pattern across an entire musical phrase, keeping the player and listener in the style of the prompt.

We further examine another dimension of style called *musical prosody*. *Prosody*, in spoken language, describes the way vocal inflections change the meaning of utterances [Shriberg et al., 1998] (e.g., "*Hellllllo*!" spoken brightly conveys a different meaning than a stern, curt "*hello.*"), despite mapping to the same vocabulary token. In music, performance dynamics and micro-improvisations on the melodic line (e.g., playing a phrase loudly, faster, slower, etc.) can have a similar prosodic effect [Heffner and Slevc, 2015]. In this demo, *Amazing Grace* offers an example of musical prosody: the player improvises micro-elements, like suspensions and rhythmic variations. Surprisingly, Aria *follows* these unseen variations, creating a novel continuation with prosodic consistency.

***Dialogue*: Harmony and Counterpoint.** Harmony and counterpoint are two ways a song weaves multiple different lines in real-time [Prout, 1891, Piston, 1947] and are seen as playing a key role in a listener's musical comprehension, by connecting the musical act to multi-speaker dialogues [Cone, 1974, Johnson, 1986, Klorman, 2016]. For *Aria* to maintain multiple voices, it is crucial for it to generate comprehensible, nuanced outputs. Liszt's *Ave Maria* transcription is notable for its multi-voice complexity: it maintains three stylistically distinct, coherent voices that the piano player weaves together: the soprano line, with far-reaching chordal leaps, the alto carrying the main melody, and the bass with a staccato grounding. *Aria* continues these three voices and, further, captures the essence of each voice in its variations: each voice continues its distinctive character and role throughout the demonstration.

***Discourse*: Phrases and Structure** Musical structure, shown both in phrasal consistency and macro-structural adherence, is an even higher-level component of sense-making [Schoenberg, 1999, Rosner, 1984]. Rosner [1984] explores *phrasal structure* in songs, drawing inspiration from linguistic parse trees developed for clause-and sentence-level analysis. Much in the same way a linguistic clause captures a single verbal idea, a musical phrase captures a single musical idea. For a song to comprehensibly express ideas to listeners, it must *introduce* and *develop* phrases. Rachmaninoff's *Piano Concerto No. 2* demonstrates *Aria*'s *phrasal* sophistication. The prompt contains a phrase taken from the first movement's secondary theme [Yang, 2025], featuring a musical arc – i.e., the primary melody weaves up and down, along with a counterpoint bass line. *Aria* develops this phrase, repeating the same arc several times through different keys, and maintaining the bass line. The idea develops without losing consistency.

According to Schoenberg [1999], a song's macro-structure or form allows audiences to understand musical phrases within broader contexts and make sense of narrative arcs. In writing, macro-structure is often studied in the context of discourse structure (e.g., essays [Montaigne, 1580], news [Van Dijk, 1988]), where structural elements have semantic meaning (e.g., *Introduction*, *Thesis*). In music, macro-structure can also be observed in two ways: (1) *semantic* awareness, or through elements which have meaning on their own (e.g., an *Overture* introduces, a *Coda* closes) [Caplin, 1998]; or (2) *formal structural repetition*, or elements which take on meaning because they occur and reoccur intentionally (e.g., $A \rightarrow B \rightarrow A$ progressions, where the A section reoccurs after the B section) [Schoenberg, 1999]. Generative models have been observed to struggle with structural coherence [Bhandari and Colton, 2024, Spangher et al., 2022, Sanchez et al., 2024] in some domains, although others have noted that different training approaches can induce more structural adherence [Tian et al., 2024].

[3]Less common with respect to the Western classical canon, which constitutes the bulk of *Aria's* training data [Bradshaw and Colton, 2025], [Bradshaw et al., 2025].

Carmen's *Overture* and Liszt's *Hungarian Rhapsody No. 6* provide us with two examples of *Aria's* semantic awareness: the *Overture* opens the demonstration while the *Rhapsody* closes it. In each, *Aria* immediately recognizes the narrative requirements posed by the prompt, and generates continuations that fulfill the narrative purpose. In the first case, *Aria* generates a slowly accelerating opening melody; in the second, it closes and ends the piece. *Spanish Song* provides us an extended example of *Aria's* formal structural awareness. The prompt presents a coherent melody, which *Aria* repeats several times in its generation. It then goes on an extended interlude before bringing the melody back, at the end of the clip. Together, the prompt and completion represent a compelling AABA structure, where the B section provides a coherent bridge between the A sections.

5 Discussion and Conclusion

We are interested in whether *Aria* is *actually* displaying these aspects of musical comprehension, or whether it is simply memorizing examples from the training dataset. While this is interesting from a scientific perspective, it also matters practically for users. Researchers have explored the effects of both *generating* novel responses and *retrieving* existing data on creative workflows [van Genugten et al., 2022, Chavula et al., 2023, Gindert and Müller, 2025]; and both have been shown to have a stimulating effect. However, *generative* approaches have been shown to aid in creativity more than *retrieval*-based approaches [Gindert and Müller, 2025, Gu et al., 2025].

In none of the completions we examined do we find evidence of memorization. Even for well-known songs, *Aria* is inventive and generates completions that, to our knowledge, are novel. To confirm that *Aria* can respond to novel input, we include in our analysis two novel compositions: the *Spanish* song written by one of the authors in 2023, and *Freygish Fino* written by Noah Smith in 2024, a professor at the University of Washington. Neither song was included in the training data. While *Aria does* seem to be less inventive in these cases, mainly playing what sounds to be improvised cadenzas, it maintains all levels of coherence that we sought to examine.

While promising, more work is required to broadly assess the impact of the *Aria-Duet* system on creativity and co-creation. Coherence, deeper musical understanding, and true generative abilities, which we have sought to explore in this work, are important characteristics for a creative assistant. So far, early results appear to be positive. The pianists we worked with were excited and enjoyed using the *Aria-Duet* system: the continuations generated were interesting and the mode of interaction was stimulating.

We acknowledge the risks of our system. Research is mixed and sharply polarized on whether modern AI-creativity tools help to stimulate and support human users. While some research points to AI tools enhancing creativity both on an individual [Doshi and Hauser, 2024] and group level [Holzner et al., 2025, Meincke et al., 2024, Lee and Chung, 2024], other studies question their capabilities [Spangher et al., 2024], with some even showing harms [Meincke et al., 2025, Kosmyna et al., 2025]. Music – despite being a fundamentally social act [Small, 1998] – faces a crisis in musician loneliness and competition [Cara et al., 2022]. While AI cannot supplant human social interaction, we hypothesize that co-creation systems have the potential to help build user skill and confidence, preserve musical forms and styles and induce greater musical participation and creation in the future. This work represents a first step in that direction, by bringing a state-of-the-art music model into a functional, usable format and demonstrating its efficacy on a Disklavier.

References

Gérard Assayag, Georges Bloch, Marc Chemillier, Arshia Cont, and Shlomo Dubnov. Omax brothers: a dynamic yopology of agents for improvization learning. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*, pages 125–132, 2006.

Keshav Bhandari and Simon Colton. Motifs, phrases, and beyond: The modelling of structure in symbolic music generation. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)*, pages 33–51. Springer, 2024.

John A. Biles. Genjam: A genetic algorithm for generating jazz solos. In *ICMC*, 1994.

Lancelot Blanchard, Perry Naseck, Stephen Brade, Kimaya Lecamwasam, Jordan Rudess, Cheng Zhi Anna Huang, and Joseph Paradiso. The jam bot, a real time system for collaborative free improvisation with music language models. In *Proceedings of the 26th International Society for Music Information Retrieval Conference*. ISMIR, 2025.

Louis Bradshaw and Simon Colton. Aria-midi: A dataset of piano midi files for symbolic music modeling. *arXiv preprint arXiv:2504.15071*, 2025.

Louis Bradshaw, Honglu Fan, Alexander Spangher, Stella Biderman, and Simon Colton. Scaling self-supervised representation learning for symbolic piano performance. *arXiv preprint arXiv:2506.23869*, 2025.

William E. Caplin. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. Oxford University Press, New York, 1998. ISBN 0-19-510480-3.

Michel A Cara, Constanza Lobos, Mario Varas, and Oscar Torres. Understanding the association between musical sophistication and well-being in music students. *International Journal of Environmental Research and Public Health*, 19(7):3867, 2022.

Catherine Chavula, Yujin Choi, and Soo Young Rieh. Searchidea: An idea generation tool to support creativity in academic search. In *Proceedings of the 2023 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '23)*, pages 161–171, 2023. doi: 10.1145/3576840.3578294.

Edward T. Cone. *The Composer's Voice*. Ernest Bloch Lectures. University of California Press, Berkeley, CA, 1974. ISBN 978-0520025080.

Mihaly Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*. Harper & Row, New York, NY, 1990.

Anil R. Doshi and Oliver P. Hauser. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28):eadn5290, 2024. doi: 10.1126/sciadv.adn5290.

Douglas Eck. Ai duet. Magenta Demo, 2017. `https://magenta.tensorflow.org/ai-duet`.

Michael Gindert and Marvin Lutz Müller. The impact of generative artificial intelligence on ideation and the performance of innovation teams (preprint), 2025. URL `https://arxiv.org/abs/2410.18357`.

Hongchao Gu, Dexun Li, Kuicai Dong, Hao Zhang, Hang Lv, Hao Wang, Defu Lian, Yong Liu, and Enhong Chen. Rapid: Efficient retrieval-augmented long text generation with writing planning and information discovery, 2025. URL `https://arxiv.org/abs/2503.00751`.

Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. Enabling factorized piano music modeling and generation with the maestro dataset. *arXiv preprint arXiv:1810.12247*, 2018.

Christopher C Heffner and L Robert Slevc. Prosodic structure as a parallel to musical structure. *Frontiers in psychology*, 6:1962, 2015.

Niklas Holzner, Sebastian Maier, and Stefan Feuerriegel. Generative AI and creativity: A systematic literature review and meta-analysis. *arXiv preprint*, 2025.

Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M Dai, Matthew D Hoffman, Monica Dinculescu, and Douglas Eck. Music transformer. *arXiv preprint arXiv:1809.04281*, 2018.

Theodore O Johnson. *An Analytical Survey of the Fifteen Sinfonias (three-part Inventions) by JS Bach*. University Press of America, 1986.

Edward Klorman. *Mozart's Music of Friends: Social Interplay in the Chamber Works*. Cambridge University Press, Cambridge, 2016. ISBN 978-1107093652.

Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, et al. Your brain on chatgpt: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv preprint*, 2025.

Susanne K Langer. *Philosophy in a new key: A study in the symbolism of reason, rite, and art*. Harvard University Press, 1942.

Byung Cheol Lee and Jaeyeon Chung. An empirical investigation of the impact of chatgpt on creativity. *Nature Human Behaviour*, 8:1906–1914, 2024. doi: 10.1038/s41562-024-01953-1.

George E. Lewis. Interacting with latter-day musical automata. *Contemporary Music Review*, 18(3): 99–112, 1999.

Lennart Meincke, Karan Girotra, Gideon Nave, Christian Terwiesch, and Karl T. Ulrich. Using large language models for idea generation in innovation. SSRN Working Paper, 2024. Wharton School Research Paper, forthcoming.

Lennart Meincke, Gideon Nave, and Christian Terwiesch. Chatgpt decreases idea diversity in brainstorming. *Nature Human Behaviour*, 2025.

Michel de Montaigne. *Essais*. Simon Millanges, Bordeaux, 1580. First edition; expanded in 1588 and posthumously in 1595.

Eugene Narmour. *The analysis and cognition of melodic complexity: The implication-realization model*. University of Chicago Press, 1992.

OpenAI. Musenet, 2019. `https://openai.com/blog/musenet`.

Francois Pachet. The continuator: Musical interaction with style. *Journal of New Music Research*, 32(3):333–341, 2003.

Walter Piston. *Counterpoint*. W. W. Norton & Company, New York, first edition, 1947. ISBN 978-0-393-09728-3.

Ebenezer Prout. *Fugue*. Library Reprints, 1891.

Burton S Rosner. A generative theory of tonal music, 1984.

Guillaume V Sanchez, Alexander Spangher, Honglu Fan, Elad Levi, and Stella Biderman. Stay on topic with classifier-free guidance. In *Proceedings of the 41st International Conference on Machine Learning*, pages 43197–43234, 2024.

Alexander Scarlatos, Yusong Wu, Ian Simon, Adam Roberts, Tim Cooijmans, Natasha Jaques, Cassie Tarakajian, and Cheng-Zhi Anna Huang. Realjam: Real-time human-ai music jamming with reinforcement learning-tuned transformers, 2025. URL `https://arxiv.org/abs/2502.21267`.

Arnold Schoenberg. *Fundamentals of Musical Composition*. Faber & Faber, London, revised ed. edition, 1999. ISBN 978-0571196586.

Arnold Schoenberg and Leonard Stein. *Style and Idea: selected writings of Arnold Schoenberg*. Univ of California Press, 1984.

Elizabeth Shriberg, Andreas Stolcke, Daniel Jurafsky, Noah Coccaro, Marie Meteer, Rebecca Bates, Paul Taylor, Klaus Ries, Rachel Martin, and Carol Van Ess-Dykema. Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and speech*, 41(3-4):443–492, 1998.

Christopher Small. *Musicking: The meanings of performing and listening*. Wesleyan University Press, 1998.

Alexander Spangher, Yao Ming, Xinyu Hua, and Nanyun Peng. Sequentially controlled text generation. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 6848–6866, 2022.

Alexander Spangher, Nanyun Peng, Sebastian Gehrmann, and Mark Dredze. Do llms plan like human writers? comparing journalist coverage of press releases with llms. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 21814–21828, 2024.

Belinda Thom. Bob: An interactive improvisational music companion. In *Proc. Autonomous Agents*, pages 309–316, 2000.

Yufei Tian, Tenghao Huang, Miri Liu, Derek Jiang, Alexander Spangher, Muhao Chen, Jonathan May, and Nanyun Peng. Are large language models capable of generating human-level narratives? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17659–17681, 2024.

Teun A Van Dijk. *News as discourse*. Routledge, 1988.

Ruben D. I. van Genugten, Roger E. Beaty, Kevin P. Madore, and Daniel L. Schacter. Does episodic retrieval contribute to creative writing? an exploratory study. *Creativity Research Journal*, 34(2): 145–158, 2022. doi: 10.1080/10400419.2021.1976451.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Enhao Yang. Musical analysis and performance practice of the first movement of rachmaninoff's piano concerto no. 2. In *Proceedings of the 2025 4th International Conference on Science Education and Art Appreciation (SEAA 2025)*, volume 952, page 300. Springer Nature, 2025.