Battle Rap as a Framework for Human-Machine Co-Creativity

Ibùkún Ọlátúnjí[1*] **Mark Sheppard**[2*]

Alma Rahat[1] Matt Jones[1] Amanda Rogers[3]

[1]Computational Foundry, Swansea University, Crymlyn Burrows, Skewen, Swansea SA10 6JW, UK
[2]University of Kent, Canterbury, Kent CT2 7NZ, UK
[3]Department of Geography, Wallace Building, Swansea University, Singleton Park, Swansea SA2 8PP, UK

Abstract

We present a human-in-the-loop GAN framework for battle rap, where a human artist (MC) serves as generator, and the AI acts as an adaptive discriminator. The AI provides feedback on rhyme complexity, coherence, and stylistic alignment, challenging the MC's improvisational skill. Fine-tuned language models emulate diverse rap styles, while voice cloning creates adversarial loops: the MC competes against stylised versions of their own voice in a dynamic, self-reflective duel. The system follows a dual-phase process: (i) an *Emulation Phase*, where AI mimics established flows to reinforce technical mastery, and (ii) an *Improvisation Phase*, where AI disrupts expectations to prompt originality. This ensures that creative growth emerges from constraint and challenge. Success is judged through MC evaluations of the AI's performance as an adversary. Framed as a study paper, this work offers a thought experiment in adversarial co-creativity, modelling how AI might inspire, rather than merely assist, human expression. Beyond computational modelling, the framework offers insights into machine-mediated creativity and how AI can be designed to provoke human creativity through improvisation, challenge, and real-time performance. The study positions the AI as a dynamic co-performer capable of eliciting novel artistic responses. As such, it contributes to emerging discourse on creative AI systems that influence, not just assist, human expression

Introduction

Rap is a primary ingredient of hip-hop music, the world's most popular genre (Lynch 2018; Texas 2015). It is a form of vocal delivery and expression that incorporates rhyme, rhythmic speech, street vernacular, and is performed over musical accompaniment (Wikipedia 2019; Edwards and Kool G Rap 2009). As Stevie Wonder describes it, '*Rap...is modern blues — a statement of how and where people are at.*' (Internet Archive 2001). We argue that rap is also a novel method for exploring human-computational creativity, as it encompasses rich linguistic, speech, musical, and socio-cultural data (Bradley 2017). Its key components are presented in Table 1.

Component	Linguistic Features	Musical Features
Content	Storytelling, metaphor, wordplay	Phonetic manipulation, assonance, consonance
Flow	Rhyme schemes, syntactic complexity	Rhythm, tempo, syncopation
Delivery	Prosody, emphasis, vocal texture	Timbre, breath control, cadence

Table 1: Key Elements of Rap

Rap contains evolving lexicons based around (i) numerous genres; and, (ii) vocal delivery based around syncopation, pitch, intonation, and cadence (Condit-Schultz 2016; Orejuela 2021). In addition, rap's short and well-documented history allows observation of the genre from its genesis to the present (Condit-Schultz 2016; Orejuela 2021).

.

Rap and Computational Creativity We propose a computational model that integrates language models and musical inputs to analyse the linguistic, stylistic, and rhythmic differences across rap genres (Copet et al. 2024; Radford et al. 2019). This approach provides insights into

*Corresponding authors: 2030349@swansea.ac.uk, ms2403@kent.ac.uk

Proceedings of the 16th International Conference on Computational Creativity 2025
ISBN: 978-989-54160-7-3

how flow[1], rhyme complexity, and thematic elements vary across styles, from battle rap to conscious hip-hop. Beyond rap, the model serves as a tool for examining broader language structures, offering insights into syntax, semantics, and linguistic creativity across different forms of expression (Akingbe and Onanuga 2018; Coscarello 2003). The model integrates constraints that reflect the MC's personal belief systems, shaping lyric generation by influencing thematic choices, and linguistic style (Edwards and Kool G Rap 2009). By embedding value-driven constraints, it examines how artistic expression is influenced by individual perspectives, cultural norms, and societal expectations (Serrano, Torres, and 2015). This provides insight into how artists navigate self-imposed or external creative limitations in music and language (Benvenga 2022).

The framework also explores the interaction between human improvisation and AI-assisted lyricism. By analysing MC responses to computational feedback in real time, it reveals how AI can influence creative decision-making, challenge artistic boundaries, and shape stylistic conventions. This deepens the understanding of AI's role as both a collaborator and an adaptive creative force in rap and other improvisational art forms (Arnold, Volzer, and Madrid 2021; Stark et al. 2023).

1. *How has rap evolved within a chosen sub-genre in terms of flow and word choice?*
   - We analyse the historical development of a specific rap sub-genre, focusing on its stylistic evolution.
   - The model allows for the integration of variable levels of artist *experiential history*, simulating how personal influences shape rap battles.
2. *How can the stylistic flow of an individual artist (B) be aligned with the broader development of their genre (A)?*
   - We introduce a framework where A and B serve as modifiers, enabling comparisons between an artist's unique flow and their genre's evolution.
   - This allows us to explore how an artist's style (B) could adapt to another genre (A2), or how a genre's conventions (A) influence a different artist (B2).
3. *What is the minimal phrase set required to define both a rap genre and an artist's individual flow?*
   - We investigate the smallest possible set of linguistic and rhythmic patterns that characterize both genre-specific and artist-specific styles.

Figure 1 describes the system in terms of the Improvisational Model and its key elements.

Battle Rap as a Creativity Game

Musical improvisation is an ideal subject for studying creativity as it involves generating novel ideas that are relevant within a given context (Csikszentmihalyi 1997; Zhang, Sjoerds, and Hommel 2020). As a specific creative act, writing verse requires an extensive vocabulary, mastery of complex rhyme patterns, and broad subject knowledge across diverse topics (Bradley 2017; Liu et al. 2012). Within rap, *freestyle*, the spontaneous composition of lyrics, is widely considered the most challenging skill to master (Edwards and Kool G Rap 2009). An fMRI study of freestyle rap revealed that it requires rapid linguistic processing, as MCs must generate meaningful, rhyming phrases in real-time while adhering to tempo and rhythm constraints (Liu et al. 2012). The process shares key characteristics with the *flow state*, where cognitive effort, heightened focus, and automaticity merge to enable fluid creative expression (Csikszentmihalyi 1997). Freestyle rap also relies on phonological awareness, lexical retrieval, and verbal dexterity (Liu et al. 2012). The ability to manipulate sound, word meanings, and rhythm in real-time is fundamental to the art form (Edwards and Kool G Rap 2009). Battle rap, a competitive format of freestyle, introduces additional constraints that amplify both its creative challenge, and its suitability for computational modelling. Unlike unstructured freestyle, battle rap requires MCs to improvise responses that directly engage with an opponent's lines while incorporating rhetorical strategies such as humor, wordplay, and personal rebuttals, all within rhythmic, structural, and/or temporal constraints (Caffeine 2020; Todd 2018). The constraints of battle rap result in two main properties:

1. *A High-Level Creativity Challenge for Human MCs* Battle rap demands that performers adapt their responses on-the-fly while maintaining coherence, linguistic complexity, and audience engagement. The interplay between structured rhythmic demands and unpredictable opponent-generated content makes it a cognitively intense and uniquely challenging form of improvisational creativity (Liu et al. 2012; Landau and Limb 2017).

2. *A Structured Framework for Computational Approaches* Battle rap's constraints and evaluative criteria (e.g.,audience reaction, lyrical complexity) allow for the development of AI systems that can generate, evaluate, and refine battle rap performances using natural language processing, machine learning, and adversarial feedback loops.

Through battle rap we can explore how AI might engage in real-time linguistic improvisation, assess human performance, and model the balance between structured constraints and creative freedom that define high-level artistic expression. This approach enables the development of computational tools capable of both augmenting human creativity, and evaluating it in a quantifiable, reproducible manner. Several computational approaches can be applied to modelling rap improvisation:

Natural Language Processing (NLP) for Lyrical Generation: State-of-the-art NLP architectures such as sequence-to-sequence models, transformers (e.g., GPT (OpenAI 2024; Xue et al. 2021) can be trained to generate context-aware rap lyrics. Unlike conventional text generation, battle rap

[1]Defined as 'all of the rhythmical and articulative features of a rapper's delivery of the lyrics' (Adams 2009)

Proceedings of the 16th International Conference on Computational Creativity 2025
ISBN: 978-989-54160-7-3

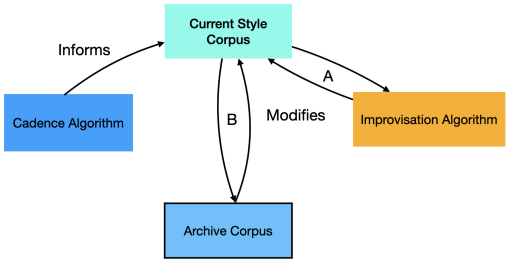

Figure 1: Improvisational Model

demands real-time linguistic agility and adversarial adaptation, requiring AI to anticipate and respond to an opponent's lyrics dynamically.

Phonological and Prosodic Analysis: As rap performance is intrinsically tied to rhythm and phonetics, computational models must incorporate phonological constraints to maintain natural flow, cadence, and prosody (Edwards and Kool G Rap 2009). Techniques such as rhyme density estimation (Hirjee and Brown 2010; Malmi et al. 2016), syllabic pattern recognition (Prinsloo and Coetzer 1990; Rogova, Demuynck, and Van Compernolle 2013), and prosody alignment (Wu et al. 2023; Cohn et al. 2021) ensure that AI-generated lyrics adhere to human-like rhythmic structures. Phonetic modelling is particularly important for rap, where multi-syllabic rhyme schemes and syncopated delivery distinguish high-level lyricism (Adams 2009).

Adversarial Learning and Response Generation: The call-and-response nature of battle rap naturally aligns with adversarial AI techniques. Generative Adversarial Networks (GANs), a machine learning framework where a generator creates data and a discriminator evaluates it in a competitive process, can be applied to text generation (Haidar et al. 2019). Generative Adversarial Imitation Learning (GAIL) (Ho and Ermon 2016) can be used to train AI to replicate expert-level rap improvisation by learning response patterns from battle data. When combined with semantic analysis and rhyme modelling, this approach could refine AI-generated counter-responses, mimicking how human MCs anticipate and react to opponent rebuttals with humour, disses, and wordplay. Reinforcement learning further optimizes response strategies by rewarding coherence, stylistic alignment, and adversarial effectiveness (Dognin et al. 2021).

Human-in-the-Loop Creativity Models: In the proposed system, Human-in-the-Loop (HITL) models keep creativity human-driven by positioning AI as an adaptive counterpart. The MC battles an AI digital twin, making the creative output the result of their interaction. Rather than replacing improvisation, the AI challenges the MC with stylistic variations and rhythmic adaptations, refining their skills through competition.

Computational Creativity Benchmarks: By analysing battle rap as a computational creativity game, we explore how AI can engage in real-time lyrical improvisation, assess human performance, and develop models that balance structured constraints with creative freedom. The ability to model adversarial language interactions has applications beyond rap, contributing to advancements in:

(i) *Dialogue systems:* Enhancing real-time conversational AI with rhetorical and adversarial elements (Chen et al. 2024).

(ii) *Stylistic text generation:* Refining personalized content creation models for creative writing (Stark et al. 2023).

(iii) *Interactive AI:* Developing AI that adapts to dynamic user input in creative and performative contexts (Gonçalo Oliveira, Mendes, and Boavida 2017) Battle rap provides a structured yet challenging test bed for evaluating AI's ability to generate, adapt, and respond in high-stakes linguistic interactions. As AI continues to develop, battle rap provides an important platform for testing human-AI improvisational collaboration and/or competition.

Related Work

This section is structured into three key research areas that directly inform AI-driven battle rap generation.

Language Generation and Speech Recognition

Language Generation: In language terms, rap can be thought of as an evolution of the poetic tradition. This view is supported by prominent poets and MCs (The Irish Times 2003; Edwards and Kool G Rap 2009). There is a long history of automated poetry systems with the earliest dating from the 18th - 19th centuries (Sharples 2015; Sharples and Perez y Perez 2022). More recently, neural network-based approaches have been used in automatic poetry classification and generation (Ghazvininejad et al. 2017; Lau et al. 2018). These approaches have also been applied to rap lyrics (Malmi et al. 2016; Xue et al. 2021). However, Transformer models have emerged as the dominant architecture for text classification, analysis, and generation tasks (Radford et al. 2019; Pichai and Hassabis 2023). While these models have shown promise in Automated Speech Recognition (ASR), their application to rap transcription remains an open challenge due to the complexities of flow, slang, and phonetic variation. Developing a battle rap system capable of generating real-time lyrical responses to a human MC requires highly accurate, low-latency ASR tailored to these linguistic and rhythmic nuances (see Figure 3)

Speech Recognition: Automated Speech Recognition (ASR) systems have made significant advancements in recent years, achieving Word Error Rates (WER) as low as 5%, approaching human transcription accuracy (Protalinski 2017; Hollands, Blackburn, and Christensen 2022; Apple 2024). Despite this progress, ASR systems still exhibit higher error rates when processing accents, dialects, and non-native English speak-

Proceedings of the 16th International Conference on Computational Creativity 2025
ISBN: 978-989-54160-7-3

ers (Wassink, Gansen, and Bartholomew 2022; Hollands, Blackburn, and Christensen 2022). These limitations are particularly relevant to the proposed system, as rap lyrics frequently incorporate non-standard English, slang, phonetic variation, and complex rhythmic structures, which pose additional challenges for ASR (Coscarello 2003; Akingbe and Onanuga 2018). Current ASR systems also perform significantly less well in speech and music,and/or singing use cases. In these situations WERs are within the 10% - 30% range (Music.AI 2024; Ou, Gu, and Wang 2022). Currently, no ASR models are specifically designed for rap transcription, presenting a gap in the research.

Prosody, Phonetics and Computational Creativity

Speech modelling: Rap presents a unique challenge for both ASR and automatic lyric generation, requiring precise syllabic stress, phoneme timing, and cadence alignment with the beat (Adams 2009; Hirjee and Brown 2010). Hirjee Brown (2009) analysed rhyme density, internal rhymes, and multi-syllabic structures in rap, while Condit-Schultz (2016) examined how MCs manipulate syllable stress to fit beats (Condit-Schultz 2016). The complexity of rap is inherently multi-modal, requiring computational models to handle both textual elements (*meaning*) and phonetic features (*sound*). As Murs explains, '*...not only do you have to make everything rhyme, but you have to add rhythm to it. Poetry doesn't have to rhyme, it just has to sound beautiful. But in rap, it has to sound beautiful, it has to be on time, and it has to rhyme*' (Edwards and Kool G Rap 2009).

While meaning and delivery are often regarded as equally important for human MCs, a case can be made that delivery plays a more crucial role. As Havoc of Mobb Deep remarked, '*...without the right flow, subject matter probably won't even matter. It's all about styles... the way you're getting your subject across. If people can't feel how you're saying it, it doesn't even matter what you're saying.*' (Wikipedia 2020; Edwards and Kool G Rap 2009) This distinction is particularly relevant for computational systems. While human MCs balance meaning and delivery, AI models may find delivery and flow more achievable, as meaning requires deeper contextual reasoning and cultural awareness. Prioritizing delivery in computational models could enable more realistic rap co-creation, even if full semantic understanding remains a challenge.

Generating semantically rich rap lyrics remains difficult, though *DeepBeat* and *DeepRapper* have explored data-driven solutions (Malmi et al. 2016; Xue et al. 2021). *DeepBeat* optimizes *rhyme density* and *coherence*, leveraging a neural network to predict next-line candidates. However, it lacks rhythm modelling, real-time lyric generation, and adversarial response capabilities - key for rap battles. *DeepRapper* builds upon this by using a Transformer to model rhyme patterns and rhythmic structures, improving fluency over DeepBeat. However, it remains a text-only system, making it unsuitable for rap battle co-creation with human MCs.

Adversarial AI & Rap Dynamics

Adversarial AI: Generative Adversarial Networks (GANs) and Reinforcement Learning (RL) have been applied to speech synthesis, music generation, and real-time improvisation, enabling AI to model stylistic variation and adaptive decision-making (Goodfellow et al. 2014; Sutton and Barto 2015). Approaches such as MuseGAN (Dong et al. 2017), SeqGAN (Yu et al. 2017), and Yi et al. (Yi et al. 2018) demonstrate AI's ability to generate structured musical compositions, poetic text, and lyrics. Learning a policy from expert demonstrations without direct interaction or reinforcement signals is challenging. Traditional methods use inverse reinforcement learning (IRL) to infer a cost function, followed by reinforcement learning (RL) to extract a policy, but this is indirect and computationally inefficient. A more effective alternative is adversarial imitation learning, where a model directly mimics expert behaviour using an approach analogous to GANs. This model-free method improves performance in complex, high-dimensional tasks. For rap improvisation, *Generative Adversarial Imitation Learning (GAIL)* provides a promising framework for AI-human co-creation (Ho and Ermon 2016). GAIL refines its policy through adversarial learning, enabling the AI to imitate expert rap performances and generate competitive, human-like responses. Table 2 summarizes the mapping of GAIL to rap battles.

Rap Dynamics: Battle rap presents a major AI challenge due to its call-and-response format, requiring low-latency NLP and speech synthesis for real-time rebuttals. Success depends on lyrical complexity, vocal tone, delivery, and crowd engagement. Battle rap's dense rhyme schemes, compound wordplay, and extended metaphors (Todd 2018) demand long-range semantic modelling for coherence. While freestyle battles may include pre-written elements, AI could employ pre-determined battle strategies, yet current text models (e.g., GPT) struggle with coherence and adversarial wordplay in real-time settings (Bender et al. 2021). Cadence, flow, and beat adaptation are equally crucial. AI-generated rap must be beat-aware and adjust flow dynamically, but most NLP models lack rhythmic and phonetic optimization. While DeepRapper improves beat alignment, generating spontaneous, flow-optimized responses remains a challenge (Xue et al. 2021). Rebuttal generation is another hurdle, as MCs rely on comedy, cultural references, and wordplay to engage the crowd. AI struggles with socio-culturally specific language, requiring context awareness and adaptability beyond current capabilities (Mirowski et al. 2024).

Proposed Method

As a proof of concept, the audio is processed from a pre-recorded source with an appropriate sample rate, though ultimately real-time data capture will be employed, enabling more realistic rap battles to take place. The audio data is extracted and passed through a cadence algorithm to evaluate the dynamics, rhythm, and structure of each phrase; creating a wire frame model of the rap. Further phrases are added to create a larger training set for the AI in order to

Proceedings of the 16th International Conference on Computational Creativity 2025
ISBN: 978-989-54160-7-3

GAIL Concept	Rap Battle Adaptation
Generator (Policy π)	The human MC improvises lyrics, responding to AI-generated lines.
Expert Demonstrations	Past battle rap performances (e.g., datasets of freestyle battles, MCs' previous performances).
Discriminator (Reward Model D)	The AI evaluates rap responses based on fluency, coherence, rhyme complexity, wordplay, and crowd engagement.
Reward Signal	AI adapts based on what makes strong battle rap responses, incorporating linguistic complexity, adversarial effectiveness, and emotional impact.

Table 2: Applying Generative Adversarial Imitation Learning (GAIL) to Rap Battles

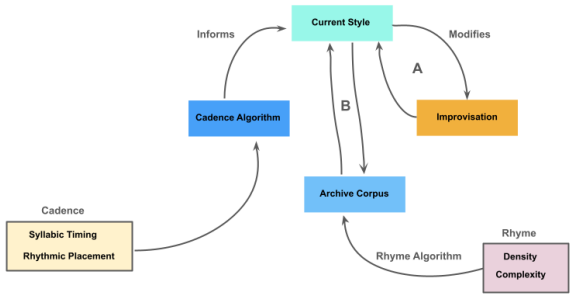

Figure 2: Simplified Cadence and Rhyme Improvisation Model

learn the style of the MC combatant. Several iterations of learning are carried out to create a more robust and detailed corpus, which is then stored as part of a wider style library. The training set can be expanded by loading library datasets from pre-recorded MCs, in order to introduce stylistic variety. Once trained, the corpus can be used to inform the AI improvisation algorithm, by generating new phrases in either the style of the human combatant or from the MC archive. In the proposed improvisational model (Figure 2), improvisation is driven by the current style corpus (A), which captures the AI's learned cadence, rhyme structures, and stylistic tendencies from prior training. This corpus forms the foundation for generating responses that align with the expected flow and delivery in a rap battle, with the system adapting to evolving stylistic cues.

Improvisation is also shaped by the archive corpus (B), which contains historical and contemporary rap data. Referencing this archive enhances lyrical complexity, cadence variation, and stylistic diversity, allowing the AI to adapt in real time. Additionally, it serves as a training resource, helping the AI model variant MC styles and flows, expanding both AI and human MC improvisational range while fostering experimentation with new delivery techniques.

Audio method

The proposed framework decouples the musicality, defined by tonality and rhythm from the language modelling process, and aims to produce a responsive and realistic battle rap model. Here pitch, cadence and amplitude data are evaluated to create a dynamic wire frame model on which to attach structured language input, reflecting the idioms and structure of a defined rap sub-genre, or stylistic approach. By integrating detailed metrics including word length, inter-word silences, and dynamic amplitude variations, the approach produces a composite cadence score that can be fine-tuned for individual phrases or entire compositions. This methodology builds on established techniques in musical feature extraction and dynamic modelling (Tzanetakis and Cook 2002; Downie 2003) extends previous work in genre-specific analysis (Condit-Schultz 2016), and draws on foundational advances in language modelling (Manning and Schütze 1999) as well as content-based music information retrieval (Casey et al. 2008; Müller 2021). This decoupling of musical analysis from the more compute intensive language modelling pipeline creates a flexible and customisable framework on which to model and expand on stylistic paradigms.

The analysis takes into account word length, inter-spacing silences, word amplitude (with respect to initial and exit volume ramping), word pitch, and beat analysis. The resultant model produces a composite cadence score, which can be calculated with varying degrees of granularity for each lyrical phrase, or as an overall score.

Both raw musical data and cadence scores can be quantified to model individual and historical genre-specific stylistic paradigms. These modelled parameters provide structured input for the improvisational algorithms (Figure 2 (A)).

Latency and Real-Time Responsiveness In assessing the effectiveness of an AI-driven rap opponent, latency between the human MC's input and the AI's response is a critical factor impacting authenticity and engagement. Three computational strategies are proposed to address latency:

1. *Cached Responses:* The AI pre-generates and stores a selection of potential responses, requiring minimal real-time computation and enabling rapid (potentially sub-second) replies. This approach closely mirrors actual

Proceedings of the 16th International Conference on Computational Creativity 2025
ISBN: 978-989-54160-7-3

practice in rap battles, where MCs often rely on memorized or pre-written material presented as spontaneous improvisation.

2. *Adaptive Rehearsal:* The AI generates initial responses beforehand but adjusts and refines them dynamically based on the MC's ongoing performance. This results in moderate latency due to additional real-time computational processing. This method achieves a practical balance between computational efficiency and improvisational authenticity.
3. *True Real-Time Generation:* Here, the AI synthesizes responses entirely in real-time, directly reacting to the MC's live input without pre-generated material. This method is computationally demanding, resulting in significantly higher latency but delivering the most authentic emulation of genuine freestyle improvisation.

Real-World vs. Computational Improvisation. Human rap battles typically feature rehearsed or semi-rehearsed content delivered with variations in emphasis and intensity, creating the illusion of spontaneous improvisation (Caffeine 2020). Therefore, the AI's utilization of prepared or adaptively adjusted content aligns closely with existing human practice (Flip Top Battles 2025).

In live rap battles, human MCs pause for 5–7 seconds between rounds, often delivering semi-rehearsed material with dynamic variation. The illusion of spontaneity is maintained through timing, emphasis, and adaptive performance (Flip Top Battles 2025). AI-driven systems that generate responses in advance or refine them adaptively are therefore not diverging from standard practice but rather emulating it closely. Key parallels include:

- Human MCs rarely freestyle entirely spontaneously; instead, they frequently utilize pre-written lyrics tailored to specific opponents or themes, performing these with nuanced variation.
- AI responses based on pre-cached or adaptively updated content reflect human norms in performance authenticity.

AI systems offer strategic advantages beyond human capability:

- They can evaluate multiple potential responses in parallel, selecting the most effective line in context.
- With reinforcement learning, these systems could improve over time—accumulating and refining a personal repertoire of adversarial tactics.

Evaluation Framework

To rigorously measure the AI opponent's effectiveness, a multi-dimensional evaluation framework is proposed, consisting of:

- *Subjective Human Assessment:* Involving self-reported evaluations from the human MC on perceived skill improvement, challenge level, and engagement, complemented by audience ratings on the AI's authenticity, creativity, and competitiveness.
- *Turing Test-Based Evaluation:* Blind evaluation by independent listeners attempting to differentiate between human and AI performances, measuring realism, stylistic consistency, and overall believability.
- *Quantitative Linguistic Metrics:* Objective analysis of linguistic and rhythmic features such as rhyme density, vocabulary diversity, rhythmic complexity, and lyrical originality to quantify changes and improvements in both human and AI performance.

Unlike conventional NLP tasks, assessing AI-generated battle rap requires both linguistic and performative evaluation. Relevant metrics include:

1. *Rhyme complexity:* measuring rhyme density and multisyllabic structures (Malmi et al. 2016).
2. *Coherence:* evaluating semantic consistency and contextual relevance
3. *Emotional impact:* For instance, detecting sentiment polarity and aggression markers in adversarial exchanges.
4. *Audience engagement:* incorporating crowd reaction prediction models.
5. *Style authenticity:* comparing AI-generated lyrics with artist-specific linguistic style (Li et al. 2024; Edwards and Kool G Rap 2009)

Developing robust battle rap evaluation metrics is critical for measuring AI's ability to mimic human-level creativity and improvisational skill. The model generates a composite cadence score by analysing the underlying rhythmic structure of the rap, which must be further evaluated for accuracy and consistency to ensure a good fit (Bigo et al. 2018; Condit-Schultz 2016). AI-generated cadence scores are compared against human-labelled benchmarks from expert MCs to establish correlation, while feature matching ensures the extracted elements, including word length, interspacing silences, amplitude, pitch, and beat alignment, align with ground truth annotations. An additional layer of verification can be elucidated by using spectral analysis comparison techniques to assess that the model accurately captures the nuances of cadence and flow (Stoica and Moses 2005).

To develop stylistic adaptability, the AI undergoes an iterative process of refinement, ensuring its output aligns with known MC styles. The wireframe model of the AI-generated rap is compared against existing MC datasets using cosine similarity on feature vectors, Dynamic Time Warping (DTW) for cadence evaluation, and Mel-Frequency Cepstral Coefficients (MFCCs) to analyse pitch and tonal properties. Transfer learning is introduced to further expand the stylistic scope, allowing the model to integrate new datasets and evolve with greater variability, generating unique and stylistically distinct improvisations. A/B testing can also be compared across different AI models, in order to identify the most effective configuration for rap improvisation.

Evaluating the quality of AI-generated phrases requires both perceptual and computational assessment. Here, evaluation methodologies can be employed by experienced MCs can provide qualitative feedback on flow, rhythmic accuracy, and stylistic authenticity. For example, a rhythmic

Proceedings of the 16th International Conference on Computational Creativity 2025
ISBN: 978-989-54160-7-3

Evaluation Dimension	Metrics and Methods
Subjective Human Assessment	Self-reported MC skill improvement, perceived challenge, and engagement; audience ratings of AI authenticity, creativity, and competitiveness.
Turing Test-Based Evaluation	Blind assessment by independent listeners evaluating realism, stylistic consistency, and believability of performances.
Quantitative Linguistic Metrics	Analysis of rhyme complexity (density, multi-syllabic structures), coherence (semantic consistency, contextual relevance), emotional impact (sentiment, aggression markers), and audience engagement prediction models.
Cadence and Flow Analysis	Composite cadence scoring using rhythmic structure features, spectral analysis, cosine similarity, Dynamic Time Warping (DTW), Mel-Frequency Cepstral Coefficients (MFCCs), and rhythmic alignment scoring.
Stylistic Authenticity and Adaptability	Feature vector comparison against MC datasets, transfer learning for stylistic diversity, expert MC qualitative feedback, and comparative A/B testing across AI configurations.
Textual Quality Assessment	Language fluency and coherence measured via BLEU scores, perplexity (PPL), and feature matching against human-labelled benchmarks.

Table 3: Multi-dimensional evaluation framework for AI-generated battle rap

alignment score can measure the fit of phrases synchronise over the underlying beat structure. Furthermore, language fluency and coherence can be assessed by comparing AI-generated text with real MC samples (BLEU scores). Additionally, Perplexity (PPL) can be employed to quantify predictability and naturalness of the generated lyric corpus, ensuring that the required framework and lyric quality is integrated into the musical framework (Papineni et al. 2002; Jelinek et al. 2005).

End-to-End System: Figure 3 illustrates the proposed AI-driven rap improvisation system, integrating voice cloning, cadence modelling, and rhyme structuring to generate adaptive lyrical responses. The human MC's speech serves as both input and feedback, guiding the AI's generative process. The AI voice clone synthesizes responses that match the MC's style, ensuring natural delivery.

The *current style module* evolves through improvisation (modifying style in real time) and historical reference (drawing from an *archive corpus*). A cadence algorithm refines rhythmic structure by analysing syllabic timing and rhythmic placement, while a rhyme algorithm evaluates rhyme density and complexity to shape lyrical flow. These components interact dynamically, with archived data reinforcing both cadence and rhyme structures to ensure stylistic coherence. A feedback loop enables continuous refinement, allowing the AI to learn from both its own output and the MC's voice, adjusting accordingly. By merging historical rap influences with real-time improvisation, the model ensures AI-generated lyrics achieve rhythmic precision, lyrical complexity, and stylistic authenticity in rap battles. Through repeated adaptation, the AI refines its freestyle capabilities, producing increasingly nuanced, context-aware responses that align more closely with human performance.

Flow Replication via TTS Models

A core system requirement is the ability to replicate human-style prosodic delivery, particularly the *flow* that exemplifies skilled MC performance. To evaluate whether current state-of-the-art text-to-speech (TTS) models meet this requirement, we conducted an experiment assessing their ability to reproduce rap-specific cadence and rhythm, using a human freestyle as reference. Flow was operationalised as the alignment between rhyme scheme and prosodic timing, with the central evaluation question framed as: does the AI-generated output *sound like* the original MC?

Two rap samples were used: (i) a human-performed freestyle baseline, and (ii) a GPT-generated text variation of the same verse. Each was synthesized using multiple TTS models and assessed on two dimensions: voice clone similarity and cadence fidelity. Evaluation combined subjective human ratings with computational metrics. This modular setup decoupled cadence replication from text generation, enabling focused analysis of a core adversarial capability in co-creative performance.Audio samples and data are available in the project repository.[2]

Human Listening Tests To capture perceptual qualities not easily quantified, human evaluators (N=8) rated each sample on voice clone similarity and flow preservation. Average ratings across all models fell between 1 (very poor) and 2 (poor), highlighting consistent failure to replicate natural cadence. These subjective scores provided a perceptual benchmark and revealed artefacts such as robotic delivery or unnatural phrasing. This experiment served as a focused probe into cadence replication in off-the-shelf TTS systems. While not a full system evaluation, it identified key barriers to adversarial responsiveness in generative rap.

Computational Analysis We developed a lightweight evaluation pipeline to quantify rhythmic similarity between original and generated audio. Mel-Frequency Cepstral Coefficients (MFCCs), a standard audio metric, captured spectral

[2] https://github.com/ArtsARKADE/ICCC25

Proceedings of the 16th International Conference on Computational Creativity 2025
ISBN: 978-989-54160-7-3

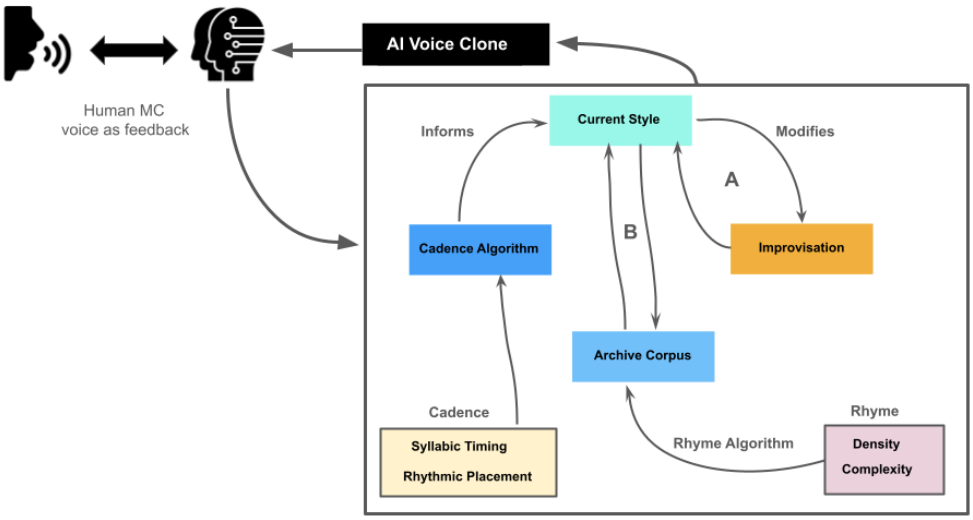

Figure 3: End-to-End Battle Rap System

features linked to perceived vocal quality. Dynamic Time Warping (DTW) assessed cadence alignment by measuring the temporal "stretch" needed to match the original performance. Higher DTW deltas indicated weaker temporal fidelity. Rhyme complexity was held constant using a GPT variation of the original text, with rhyme density calculations confirming phonemic similarity. A 3D bar chart summarises the MFCC and DTW metrics across TTS models (Figure 4).

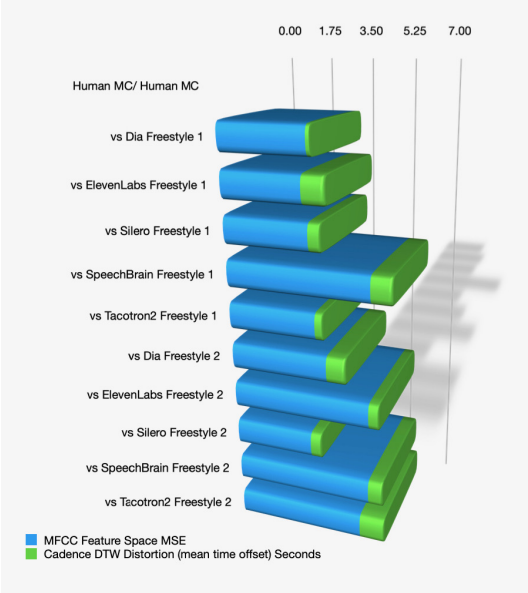

Figure 4: Human evaluator scores of TTS rap outputs across models and samples. Full data and code available in GitHub repository.

Rationale for Dual Evaluation For an AI to function as a creative adversary, provoking new human responses rather than imitating past styles, it must exert fine-grained control over cadence. Perceptual tests alone often miss subtle timing mismatches that limit this capacity. By combining human ratings with MFCC-based and cadence alignment metrics, we gained structural and perceptual insight. Together, these evaluations expose limitations in adversarial responsiveness and shape the design priorities discussed in the next section.

Discussion and Future Work

This paper explores AI's potential as both collaborator and competitor within rap performance. Several areas invite further investigation, including real-time responsiveness, the balance between rehearsed and improvised content, and the implications of AI surpassing human artistic abilities.

Discussion The proposed model presents a structured approach to generating AI-driven rap improvisation, leveraging a composite cadence score to quantify rhythmic structure and combining it with realistic lyrical content generated from large language models (LLMs). The model captures stylistic nuances across differing MC styles and is enhanced by iterative training and transfer learning, enabling the system to produce varied and authentic improvisations. However, several challenges remain in order to fully automate high-quality AI rap generation while remaining true to the genre's historical and expressive roots. The subjective nature of flow, authenticity, and lyrical coherence complicates evaluation, necessitating both computational metrics and human judgement. Real-time performance constraints are important for maintaining computational efficiency. At present, the model focuses on an offline approach to prioritise stylistic accuracy and improvisational relevance while

Proceedings of the 16th International Conference on Computational Creativity 2025
ISBN: 978-989-54160-7-3

balancing creative autonomy and learned structure. While an AI can generate stylistically appropriate phrases, it remains constrained by its training corpus, raising questions about copyright and authenticity with regard to AI-generated output that closely resembles human artists' styles. Given that hip-hop is deeply rooted both as a lived experience and as an evolving cultural history, AI models must be trained responsibly and with respect to the art form, preserving its integrity and social context.

The battle rap improvisational model has broader applications in linguistic AI, speech synthesis, automated storytelling, and conversational AI. Its approach to analysing cadence, rhythm, and flow may also enhance natural language processing (NLP) models, leading to improvements in speech recognition, voice assistants, and real-time dialogue generation.

Optimising real-time processing is crucial for freestyle battles, requiring efficient feature extraction, phrase generation, and beat synchronization to minimize latency. Future work could explore low-latency deep learning models to reduce computational overhead while preserving performance. Leveraging additional MC datasets, transfer learning, and fine-tuning could improve AI responsiveness by integrating historical and contemporary rap influences. Such improvements would not only enhance stylistic fidelity, but also support the real-time demands of adversarial performance contexts. Future research may also refine cadence analysis techniques by incorporating prosodic modelling and sentiment-aware lyric generation, improving authenticity and coherence in AI-generated verses. Advancing AI-driven rap battles represents a step toward machines engaging in real-time, creative, and adversarial discourse—key components of human intelligence found in debate, negotiation, and artistic improvisation (Guilford 1967; Boden 1992). This research contributes to a broader effort in computational creativity, demonstrating how AI can learn to navigate spontaneity, stylistic nuance, and competitive interaction, with implications extending beyond music to conversational AI, linguistic modelling, and real-time creative collaboration.

AI models trained in rap improvisation could eventually serve as intelligent songwriting assistants, helping artists refine their flows, experiment with rhythmic structures, and generate new lyrical ideas. This opens the possibility of live AI performers dynamically integrated into concert environments. Extending this to virtual reality (VR) performances or gaming experiences could be a logical next step, where AI-generated rap dynamically responds to audience input or player actions. Our approach contributes to the broader field of computational creativity, showing how deep learning models can be fine-tuned to produce expressive, human-like artistic outputs and potentially innovate and evolve new stylistic paradigms. This raises important questions about whether AI can eventually surpass mimicry and develop its own distinct artistic identity (Du Sautoy 2020).

Implications of AI Surpassing Human Artistic Abilities AI-driven rap improvisation has wider implications across music, AI creativity, human-AI collaboration, and computational linguistics. Developing new improvisational tools for musicians and producers, such as virtual compositional assistants, could significantly enhance human-AI collaboration in creative industries. The potential scenario of an AI surpassing human MC skill levels raises significant philosophical and creative questions: (i) Might an AI that exceeds human creative performance stimulate increased artistic ambition by challenging human artists to evolve and innovate? (ii) Or could it demotivate artists by creating perceptions of redundancy, potentially diminishing the perceived value of human creative expression? These considerations underline the need for further investigation into the sociocultural implications of AI's evolving role in creative practices beyond rap battles, particularly in how AI-human interactions reshape artistic identity, motivation, and creativity itself.

Creative Adversarial Evaluation and Future Work The proposed system imagines AI not as a co-pilot, but as a sparring partner that pushes the boundaries of what it means to create. This leads to a model for AI evaluation that we refer to as a *Creative Adversarial Turing Test* (CATT). Unlike traditional Turing Tests (Turing 1950), which reward mimicry, a CATT assesses whether an AI can provoke a human to create at a higher level. Within the rap battle context, the AI is not a passive tool but a rival challenging the MC through feedback on rhyme, flow, and stylistic coherence. The provocation triggers a creative dialogue, prompting the human to iterate in response to the AI's performance. This model envisages AI evaluation as a test of *stimulation*, not *deception*. The outcome is not judged by whether the AI *sounds human*, but whether the *human sounds more inspired*. This adversarial framing reflects how creative tension, not comfort, often drives innovation (pp. 221–222, Du Sautoy 2020). CATT offers a test bed for exploring human adaptability under machine provocation. Beyond rap, it invites new forms of human-AI dialogue in fields such as education, design, and performance, where improvisation shapes meaning in real-time. The study re-frames the question from "Can AI create? to "Can AI compel *us* to create differently?"

This paper lays the groundwork by measuring flow preservation and adversarial response. These early results demonstrate what today's models lack, and why adversarial challenge might be a way drive progress in co-creative AI. This provides an initial step toward modelling the dynamic tension that underpins the Creative Adversarial Turing Test. Over time, we aim to develop structured CATT benchmarks that assess co-evolution: how well an AI can stretch, disrupt, or sharpen human creativity through tension. It challenges designers to think beyond assistive tools toward performative, even confrontational, systems. This direction raises new questions about authorship, creative ownership, and the emotional stakes of creativity when it emerges from competition rather than collaboration. Ultimately, the CATT framework imagines AI as an active agent in artistic development, provoking growth rather than imitating past success.

Proceedings of the 16th International Conference on Computational Creativity 2025
ISBN: 978-989-54160-7-3

References

Adams, K. 2009. On the metrical techniques of flow in rap music. *Music Theory Online* 15.

Akingbe, N., and Onanuga, P. A. 2018. Leveraging poetry on the airwaves: Appropriating linguistic creativity in nigerian hip hop lyrics. *Journal of the Musical Arts in Africa* 15:19–40.

Apple. 2024. Humanizing word error rate for asr transcript readability and accessibility.

Arnold, K.; Volzer, A.; and Madrid, N. 2021. Generative models can help writers without writing for them. In *Joint Proceedings of the ACM IUI 2021 Workshops*.

Bender, E. M.; Gebru, T.; McMillan-Major, A.; and Shmitchell, S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 610–623. New York, NY, USA: Association for Computing Machinery.

Benvenga, L. 2022. Hip-hop, identity, and conflict: Practices and transformations of a metropolitan culture. *Frontiers in Sociology* 7.

Bigo, L.; Feisthauer, L.; Giraud, M.; and Levé, F. 2018. Relevance of musical features for cadence detection. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*.

Boden, M. 1992. *The Creative Mind*. London: Abacus.

Bradley, A. 2017. *Book of Rhymes: The Poetics of Hip Hop*. Basic Civitas.

Caffeine. 2020. A crash course on battle rap and the ultimate rap league.

Casey, M.; Veltkamp, R.; Goto, M.; Leman, M.; Rhodes, C.; and Slaney, M. 2008. Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE* 96:668–696.

Chen, K.; Shao, A.; Burapacheep, J.; and Li, Y. 2024. Conversational ai and equity through assessing gpt-3's communication with diverse social groups on contentious topics. *Nature Scientific Reports* 14:1561.

Cohn, M.; Predeck, K.; Sarian, M.; and Zellou, G. 2021. Prosodic alignment toward emotionally expressive speech: Comparing human and alexa model talkers. *Speech Communication* 135:66–75.

Condit-Schultz, N. 2016. *MCFlow: A Digital Corpus of Rap Flow*. Ph.D. Dissertation, Ohio State University.

Copet, J.; Kreuk, F.; Gat, I.; Remez, T.; Kant, D.; Synnaeve, G.; Adi, Y.; and Défossez, A. 2024. Simple and Controllable Music Generation.

Coscarello, C. 2003. *The Word Out : A Stylistic Analysis of Rap Music*. Ph.D. Dissertation, Montclair State University.

Csikszentmihalyi, M. 1997. *Creativity: Flow and the Psychology of Discovery and Invention*. Harper Perennial.

Dognin, P.; Padhi, I.; Melnyk, I.; and Das, P. 2021. Regen: Reinforcement learning for text and knowledge base generation using pretrained language models. In Moens, M.-F.; Huang, X.; Specia, L.; and Yih, S. W.-t., eds., *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 1084–1099. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.

Dong, H.-W.; Hsiao, W.-Y.; Yang, L.-C.; and Yang, Y.-H. 2017. Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment.

Downie, J. 2003. Music information retrieval. *Annual Review of Information Science and Technology* 37:295–340.

Du Sautoy, M. 2020. *The creativity code : how AI is learning to write, paint and think*. 4Th Estate.

Edwards, P., and Kool G Rap. 2009. *How to Rap : The Art and Science of the Hip-Hop MC*. Chicago Review Press.

Flip Top Battles. 2025. Flip top battle league.

Ghazvininejad, M.; Shi, X.; Priyadarshi, J.; and Knight, K. 2017. Hafez: an Interactive Poetry Generation System. 43–48.

Gonçalo Oliveira, H.; Mendes, T.; and Boavida, A. 2017. Co-poetryme: a co-creative interface for the composition of poetry. *Proceedings of the 10th International Conference on Natural Language Generation*.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Nets.

Guilford, J. P. 1967. The nature of human intelligence.

Haidar, M. A.; Rezagholizadeh, M.; Omri, A. D.; and Rashid, A. 2019. Latent code and text-based generative adversarial networks for soft-text generation. In Burstein, J.; Doran, C.; and Solorio, T., eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics.

Hirjee, H., and Brown, D. 2010. Using Automated Rhyme Detection to Characterize Rhyming Style in Rap Music. *Empirical Musicology Review* 5:121–145.

Ho, J., and Ermon, S. 2016. Generative Adversarial Imitation Learning.

Hollands, S.; Blackburn, D.; and Christensen, H. 2022. Evaluating the performance of state-of-the-art asr systems on non-native english using corpora with extensive language background variation. In *Proceedings of the Annual Conference of the International Speech Communication Association. Interspeech 2022, 18-22 Sep 2022, Incheon, Korea*, 3958–3962. Interspeech 2022.

Internet Archive. 2001. Mtv.com - news -madonna, elton, stevie wonder defend eminem.

Jelinek, F.; Mercer, R. L.; Bahl, L. R.; and Baker, J. K. 2005. Perplexity—a measure of the difficulty of speech recognition tasks. *The Journal of the Acoustical Society of America* 62(S1):S63–S63.

Landau, A. T., and Limb, C. J. 2017. The neuroscience of improvisation. *Music Educators Journal* 103(3):27–33.

ISBN: 978-989-54160-7-3

Lau, J. H.; Cohn, T.; Baldwin, T.; Brooke, J.; and Hammond, A. 2018. Deep-speare: A joint neural model of poetic language, meter and rhyme. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Li, Y. A.; Jiang, X.; Darefsky, J.; Zhu, G.; and Mesgarani, N. 2024. Style-talker: Finetuning audio language model and style-based text-to-speech model for fast spoken dialogue generation.

Liu, S.; Chow, H. M.; Xu, Y.; Erkkinen, M. G.; Swett, K. E.; Eagle, M. W.; Rizik-Baer, D. A.; and Braun, A. R. 2012. Neural Correlates of Lyrical Improvisation: An fMRI Study of Freestyle Rap. *Scientific Reports* 2.

Lynch, J. 2018. Hip-hop passes rock to become most popular music genre for first time in history: Nielsen.

Malmi, E.; Takala, P.; Toivonen, H.; Raiko, T.; and Gionis, A. 2016. DopeLearning: A Computational Approach to Rap Lyrics Generation. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Manning, C. D., and Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press.

Mirowski, P. W.; Love, J.; Mathewson, K. W.; and Mohamed, S. 2024. A robot walks into a bar: Can language models serve as creativity support tools for comedy? an evaluation of llms' humour alignment with comedians.

Music.AI. 2024. Comparative analysis shows music.ai outperforms openai in lyric transcription accuracy.

Müller, M. 2021. *Fundamentals of Music Processing: Using Python and Jupyter Notebooks*. Springer International Publishing.

OpenAI. 2024. ChatGPT Overview.

Orejuela, F. 2021. History of Rap Hip-Hop.

Ou, L.; Gu, X.; and Wang, Y. 2022. Transfer learning of wav2vec 2.0 for automatic lyric transcription.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. Bleu: a method for automatic evaluation of machine translation. In Isabelle, P.; Charniak, E.; and Lin, D., eds., *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics.

Pichai, S., and Hassabis, D. 2023. Introducing gemini: our largest and most capable ai model.

Prinsloo, G., and Coetzer, M. 1990. Automatic syllabification and phoneme class labelling with a phonologically based hidden markov model and adaptive acoustical features. *Computer Speech Language* 4(3):247–262.

Protalinski, E. 2017. Google's speech recognition technology now has a 4.9% word error rate.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language Models are Unsupervised Multitask Learners.

Rogova, K.; Demuynck, K.; and Van Compernolle, D. 2013. Automatic syllabification using segmental conditional random fields. *Computational Linguistics in the Netherlands Journal* 3:34–48.

Serrano, S.; Torres, A.; and , I. T. 2015. *The Rap Yearbook : The Most Important Rap Song from Every Year Since 1979, Discussed, Debated, and Deconstructed*. Abrams Image.

Sharples, M., and Perez y Perez, R. 2022. *Story Machines : How Computers Have Become Creative Writers*. Routledge, first edition.

Sharples, M. 2015. John Clark's Latin Verse Machine: 19th Century Computational Creativity.

Stark, J.; Tang, A.; Park, J.; and Wigdor, D. 2023. Can ai support fiction writers without writing for them?

Stoica, P., and Moses, R. 2005. Spectral analysis of signals. *Prentice Hall* 1–447.

Sutton, R. S., and Barto, A. 2015. Reinforcement Learning : An Introduction.

Texas, A. 2015. Hip-hop is the most listened to genre in the world.

The Irish Times. 2003. Heaney lauds 'verbal energy' of Eminem.

Todd, B. 2018. Rapper and poet mark grist gives tips for rap battle global domination.

Turing, A. 1950. Computing machinery and intelligence. *Mind* 59:433–460.

Tzanetakis, G., and Cook, P. 2002. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing* 10(5):293–302.

Wassink, A. B.; Gansen, C.; and Bartholomew, I. 2022. Uneven success: automatic speech recognition and ethnicity-related dialects. *Speech Communication* 140:50–70.

Wikipedia. 2019. Rapping.

Wikipedia. 2020. Mobb Deep.

Wu, H.; Yun, J.; Li, X.; Huang, H.; and Liu, C. 2023. Using a forced aligner for prosody research. *Humanities and Social Sciences Communications* 10.

Xue, L.; Song, K.; Wu, D.; Tan, X.; Zhang, N.; Qin, T.; Zhang, W.-Q.; and Liu, T.-Y. 2021. DeepRapper: Neural rap generation with rhyme and rhythm modeling. 69–81.

Yi, X.; Sun, M.; Li, R.; and Li, W. 2018. Automatic Poetry Generation with Mutual Reinforcement Learning. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

Yu, L.; Zhang, W.; Wang, J.; and Yu, Y. 2017. Seqgan: Sequence generative adversarial nets with policy gradient. *Proceedings of the AAAI Conference on Artificial Intelligence* 31.

Zhang, W.; Sjoerds, Z.; and Hommel, B. 2020. Metacontrol of human creativity: The neurocognitive mechanisms of convergent and divergent thinking. *NeuroImage* 210.

Proceedings of the 16th International Conference on Computational Creativity 2025
ISBN: 978-989-54160-7-3