Balancing Indeterminacy and Structure: Neural Text Generation for Artistic Inspiration

Olga Vechtomova[0000−0001−7371−0837] and Gaurav Sahu[0000−0002−9950−6313]

University of Waterloo, Ontario, Canada
{ovechtom, gsahu}@uwaterloo.ca

Abstract. This paper explores the potential of neural text generative models to produce poetic lines that can serve as seeds for artistic inspiration. Drawing on theories of aesthetic perception and creativity, we hypothesize that lines characterized by indeterminacy and occupying an intermediate space between randomness and predictability are most effective at inspiring creative work. We compare two architectures: Long Short-Term Memory Variational Autoencoders (LSTM-VAE) and Transformer-based Large Language Models (LLMs), analyzing their outputs using metrics that capture linguistic, stylistic, and poetic qualities. Our analysis reveals that LSTM-VAE generates lines with higher global entropy and more variable syntactic patterns while maintaining lower levels of pretentiousness compared to LLMs. While LLM-generated lines demonstrate richer conventional poetic imagery, they often present overly finished forms that may inhibit rather than stimulate creative exploration. The study's findings suggest that smaller, more focused models trained on curated datasets might be more effective at generating semantically open lines than larger, general-purpose language models. This is particularly relevant during the Seed phase of creativity, where the goal is not to produce polished artistic output but to help artists enter a state of heightened perception and creative possibility. Our work contributes to understanding how computational systems can effectively support human creativity while providing a framework for evaluating AI systems designed to bring artists into the creative state.

Keywords: Computational creativity · Artistic inspiration · Neural text generation · Aesthetic perception · Creative process.

1 Introduction

Artistic inspiration is a complex and often elusive phenomenon that is essential to the creative process yet challenging to systematically evoke or understand. Despite significant advances in AI-generated content, little research has systematically investigated how different neural architectures might serve as tools for inspiring human creativity rather than simply generating finished artistic works. This gap is particularly notable given the increasing integration of AI tools in creative processes.

In the context of the Seed phase of creativity, where initial ideas and stimuli give rise to artistic work [17], various techniques like the cut-up have been employed to bring the artist into the creative state. With the advances in neural text generative models, new possibilities have emerged for generating stimuli such as short poetic and lyric lines that can serve as seeds for inspiration.

The goal of these stimuli is not to push fully formed ideas to the artist by providing ready-made poems or lines that the artist would use directly, but to pull the artist into the state of inspiration where creating art becomes a possibility. This is a state that is commonly associated with heightened attention and perception [15,26]. Furthermore, creating genuine art and poetry often requires the ability to see things in a new "strange" way – the concept of "ostranenie" or estrangement formulated by Shklovsky [21].

Indeterminate and surprising stimuli, such as abstract patterns in the visual domain or unresolved lines with semantic openness and syntactic defamiliarization in text can help artists enter this state. Leonardo Da Vinci advised young artists to look at old stained walls and streaked stones to get inspiration [22]. Indeterminate stimuli like these are projective and have open spaces that the artists fill with their own ideas, interpretations and images. The meaning inherent in these stimuli is unresolved. Instead, the artist forms the meaning as the stimulus is processed by the mind [15,26]. Furthermore, the process of the mind trying to consciously or subconsciously resolve these indeterminate stimuli is bringing the artist into the creative or inspired state with heightened perceptual acuity.

According to research in the psychology of aesthetics, people tend to engage more with the stimuli of intermediate complexity – the phenomenon described as the inverted U-shaped (Wundt) effect [3,6]. In information-theoretic terms, the complexity-simplicity continuum can be viewed as randomness-predictability.

Following art and aesthetics theories, as well as artists' observations, intuitions, and artistic practices, we hypothesize that generated lines of intermediate complexity are the most conducive to bringing the artist into the inspired state during the Seed phase of the creative process. However, the simplicity-complexity (or predictability-randomness) continuum is too coarse-grained when applied to text. We need to understand what characteristics the lines occupying the midrange region possess.

To address these questions, we focus on two neural network architectures: Long Short-Term Memory Variational Autoencoders (LSTM-VAE) and Transformer based Large Language Models (LLMs). We chose these architectures because they represent distinct approaches to text generation: LSTM-VAE's more constrained, focused generation versus LLMs' broader, context-aware approach. While LLMs have demonstrated remarkable capabilities in generating coherent and contextually appropriate long-form text, we hypothesize that LSTM-VAE models may be better suited for producing surprising and evocative sentence-length lines that occupy the intermediate space between randomness and predictability.

The contributions of this paper are threefold:

1. We discuss the role of randomness, predictability, indeterminacy and estrangement in art and literature. By exploring art theories as well as historical and contemporary artistic techniques that apply these concepts, we establish a foundation for understanding how generative models can contribute to the Seed phase of creativity.
2. We compare two neural generative architectures: LSTM-VAE and Transformer-based LLM, in the context of text generation for artistic inspiration. By examining their respective training objectives, data requirements, and generative capacities, we discuss the suitability of each model for producing lines that occupy the productive mid-range between predictability (simplicity) and randomness (complexity).
3. We perform comparative analysis of lines generated by each model using automatic metrics that measure linguistic and stylistic properties, including entropy, surprisal, abstractedness, valency, syntactic complexity, pretentiousness, simplicity and indeterminacy.

While our focus is specifically on text generation for poetic and lyrical inspiration, our findings have broader implications for understanding how AI systems can support rather than replace human creativity.

2 Inspiration and the Creative Process

Artistic inspiration remains one of the least understood aspects of human experience. According to the 19th century American artist Robert Henri, "the object isn't to make art, it's to be in that wonderful state which makes art inevitable." Rubin (2023) outlined three stages of the creative process:

- **Seed phase**. We are completely open to any stimulus that might lead us to that state of mind where art becomes inevitable: thoughts, sensory inputs, overheard conversations, or unusual phrases. We don't get to choose when a noticing or inspiration comes. We can only be there to receive it." [17]
- **Experimentation**. Ideas and images from the Seed phase coalesce into new forms as we explore possibilities while engaging our critical judgment.
- **Crafting**. The final stage where a clear sense of direction has arisen" [17] and we apply technical mastery to realize our artistic vision.

Our study focuses on the Seed phase. The ephemeral nature of inspiration makes it particularly challenging to study—these moments are unique and unrepeatable, striking from unexpected sources like fleeting memories, sounds, or scents. The question then becomes: how can artists actively cultivate these moments of inspiration rather than passively waiting for them? The next section explores various techniques artists have developed to summon this elusive state.

3 Randomness and Predictability in Art and Literature

Many modern artists and writers use the concepts of randomness and predictability with various artistic and literary intents. In this section we will discuss creative techniques, art movements and theories that explored these concepts.

One technique commonly used by poets and artists to summon inspiration is the cut-up technique. Proposed by Tristan Tzara and other Dada artists in the early 20th century, it involved using scissors to cut up printed text into words and letting them fall in random order. While Dadaists intended this as an anti-art statement, William Burroughs later popularized it as a literary device. Burroughs argued that language is a "lock" that restricts creativity and confines us to predictable patterns [4]. Music artists David Bowie [9] and Kurt Cobain [5] used this technique in their songwriting. Bowie described using random juxtaposition of concepts either directly or as stimulation for new lyric ideas, saying: "What I've used it for, more than anything else, is igniting anything that might be in my imagination." [2].

Another related technique is syntactic defamiliarization. Russian Formalist Viktor Shklovsky's theoretical framework "ostranenie" (defamiliarization or estrangement) shows how breaking conventional syntactic-semantic relationships can make language feel fresh and strange [21]. When artists disrupt familiar syntactic structures, they force readers to engage with language more fully and form their own meanings.

Many 20th century visual artists intentionally incorporated randomness in their work, including Jean Arp, Georg Nees, Kurt Schwitters and Vera Molnar. Nees, considered the author of the first computer art, stated that randomness alone cannot create art - randomness without aesthetic structure is simply chaos. There must be structure to narrow down random outcomes and produce discernible aesthetic information [11]. Molnar observed that computer programs using random generators help explore more possibilities in finding artistic forms than otherwise possible [14].

Kurt Schwitters used randomness and surprise in his Merz-poems through collage, incorporating varied textual elements from different sources. Writing about his poem "An Anna Blume", he noted: "'I love thou' is more expressive than 'I love thee', because it's unusual. Had I written 'I love thee' nobody would have noticed that I love Anna Blume" [19].

American novelist Tom Spanbauer's 'burnt tongue' technique deliberately misuses language to make it fresh and engaging. As he puts it: "As writers, part of your job is to bend and misuse your words and force readers to read as if they've just learned how." His student Chuck Palahniuk describes it as "saying something wrong, twisting it to slow down the reader." Spanbauer's intuition is validated by eyetracking studies showing that predictable words are skipped more often and elicit shorter fixations [13].

Conversely, some artists deliberately employed predictability as a creative device. Pop-art of the 1950s and 1960s exemplifies this approach, using recognizable imagery and cliché statements to critique consumer culture, as seen in Barbara Kruger's works with slogans like "I shop therefore I am."[12]. Conceptual artist Jenny Holzer intentionally used banal expressions in her work "Truisms", with trivial-sounding statements like "A little knowledge goes a long way" that challenge viewers' perceptions of art and language.[8]

4 Complexity, Indeterminacy and Aesthetic Perception

The relationship between aesthetic pleasure and complexity, known as the inverted U-shaped or "Wundt" effect, suggests that stimuli of intermediate complexity are most pleasing. This effect, first explored by Berlyne [3], has been observed in various domains such as music [6] and product design [1].

Complexity is typically expressed in features such as predictability and surprise. Gold et al [6] offered a learning-based explanation for the inverted U-shaped effect. When events are completely predictable or fully random, little to no learning occurs as patterns are either trivial or impossible to discern. However, events with intermediate predictability enhance learning and meaningful engagement with the stimulus. While this effect appears across aesthetic domains, evidence suggests that our liking of aesthetic stimuli remains highly individual.

Indeterminacy is another important characteristic of aesthetic stimuli. The meaning of indeterminate artifacts (text lines, sounds, or images) emerges only through the perceiver's internalization process. For example, in visual art, indeterminacy manifests when artworks invite multiple interpretations. As Pepperell noted, resolving indeterminate images creates heightened perception and attention as our habitual recognition processes are suspended [15]. This heightened state can itself lead to inspiration.

We hypothesize that AI-generated poetic lines in the mid-region of the random-predictable spectrum are most likely to be useful as seeds for inspiration. In the rest of the paper, we explore two neural text generative models' potential to generate lines that occupy this region. We begin by introducing these models before comparing their outputs as potential sources of seed lines.

5 Neural Text Generative Models

Text generation has seen two major breakthroughs: the adoption of Long Short Term Memory (LSTM) networks in 2016 [7], and the introduction of Transformer-based Large Language Models in the early 2020s [23]. These generative models produce original sequences—texts that are not directly derived from inputs through replication or rule-based responses—with or without conditioning signals such as prompts, images, or audio.

This section examines two neural network architectures: Variational Autoencoders based on Long-Short Term Memory Networks (LSTM-VAE) and Transformer-based Large Language Models (LLMs).

LSTM-VAE. The Variational Autoencoder (VAE) [10] is a stochastic neural generative model with an encoder-decoder architecture. The VAE training objective combines reconstruction loss with Kullback-Leibler divergence:

$$J_{\text{VAE}} = \underbrace{-\sum_{t=1}^{n} \log p(x_i|\boldsymbol{z}, x_1 \cdots x_{i-1})}_{\text{reconstruction loss}} + \underbrace{\text{KL}\left(q\phi(\boldsymbol{z}|x), ||, p(\boldsymbol{z})\right)}_{\text{KL divergence}} \tag{1}$$

where ϕ and θ are encoder and decoder parameters, and the KL term encourages the posterior distribution to be close to a prior (typically standard normal).

After training, text generation involves sampling a latent vector $\boldsymbol{z}$ from the prior and passing it through the decoder. The model can be extended to Conditional VAE (CVAE) by incorporating additional signals like audio [25]. LSTM-VAE architectures excel at generating sentence-length texts.

Large Language Models (LLMs). Large Language Models (LLMs) based on the Transformer architecture represent the current state-of-the-art in text generation. Their core component, the self-attention mechanism, processes all words in a sequence simultaneously, enabling effective modeling of long-range dependencies. LLMs are trained to predict the next word in a sequence by minimizing the negative log-likelihood:

$$J_{\text{LLM}} = -\sum_{i=1}^{n} \log p(x_i|x_1, \ldots, x_{i-1}) \tag{2}$$

where x_i is the i-th word and $p(x_i|x_1, \ldots, x_{i-1})$ represents the probability of the next word given its context. Unlike LSTM-VAE's sentence-level focus, trained LLMs can generate longer, contextually coherent texts through sequential word sampling.

6 AI-generated Lines as Sources of Inspiration During the Seed Phase of the Creative Process

Transformer-based LLMs are more suited for generating long-form texts and also attending to longer contexts, compared to VAE architectures. The VAE, on the other hand, generates short sentence-level lines of text. During the Seed phase of the creative process, we may not necessarily need long-form texts to inspire us, as they can be too close-ended. In contrast, short snippets of ideas and incomplete expressions are arguably more useful as they are open to interpretation.

Poetic lines that spark creativity tend to be indeterminate and open-ended rather than presenting finished thoughts. By resisting fixed interpretation, they invite readers to explore meanings beyond the surface, suggesting images and ideas that the artistic mind can develop further.

The reason why some of the most innovative and avant-garde artists used the cut-up technique is because they wanted to be surprised and led down the creative path that they couldn't have possibly discovered. The cut-up technique, however, is pure chaos. As the inverted U-shape hypothesis suggests, the more interesting results could be somewhere between pure randomness and complete predictability.

We hypothesize that LSTM-VAE and its conditional variants are more likely to produce lines that possess such qualities than LLM. Below we outline the reasons for this hypothesis:

Small curated dataset. LSTM-VAE requires only a small fraction of the training data compared to LLMs - in our work, we used only 20K lyric lines to train a VAE, while LLMs require billions of texts. This difference has two important implications. First, with VAE, we can carefully curate our dataset to include only high-quality texts that reflect specific artistic preferences. Second, current LLMs, trained on essentially all available texts from the Internet, lack this curation. This affects their generated poetic or lyric texts, which often trend toward being generic and less original in both thought and form.

Training objective. Both VAE reconstruction loss and LLM loss are autoregressive, computing conditional probability of the next token based on previous tokens in the sequence. However, VAE reconstruction loss additionally incorporates the latent vector z - a crucial distinction, as VAE learns to model the data as latent variable space, while LLM doesn't. This latent vector, sampled randomly from the latent space, encodes the semantics and syntax of a hypothetical sentence. The VAE latent space is smooth, meaning that sampling from a random point in the 128-dimensional latent space rarely generates sentences seen during training. Our experiments demonstrate that about 97.5% of generated lines are original (not copies from the training set), and 99% are unique, showing the VAE's capacity for diverse, non-repetitive output [24]. This diversity stems from the VAE decoder conditioning each word prediction on both the preceding tokens and the latent vector z. The model accounts for the z vector's signal, perturbing the autoregressive prediction and leading to less predictable sequences compared to LLMs.

Model size. Modern LLMs contain billions of trainable parameters and use self-attention mechanisms with multiple attention heads, allowing them to predict next words with remarkable accuracy based on context. LSTM-VAE, by contrast, has only hundreds of thousands of parameters. This limitation actually becomes an advantage for creative generation - with less capacity to memorize the training set, the model produces more surprising and unpredictable word sequences.

7 Metrics for Assessing Inspirational Value

We compiled a dataset, which comprises lyric lines generated by two systems: an LSTM-VAE-based system trained on lyrics in the Rock genre, which generated lyric lines conditioned on MEL-spectrograms of instrumental music [25], and ChatGPT-4o, using a prompt that was specifically designed to get the model to generate lyric lines that would be useful in the Seed stage of the creative process[1]. The dataset contains 2000 lines in total, with an equal number of lines from each system.

Drawing on the U-shaped effect hypothesis [3,6] that suggests peak aesthetic engagement at intermediate complexity levels, we developed metrics to identify

[1] https://sites.google.com/view/balancing-indeterminacy/paper (supplementary material)

lines that occupy this productive middle ground. Following [18], we analyzed both objective linguistic features and interpretive poetic qualities.

Our analysis encompasses ten features: four objective metrics (surprisal, contextual entropy, global entropy, and syntactic complexity) measuring linguistic structure, and six interpretive measures (imagery, abstractedness, valence, indeterminacy, simplicity, and pretentiousness) evaluating poetic qualities. The interpretive measures were obtained using LLaMA-3-70b in a zero-shot manner, rating each line on a 5-point scale and computing weighted scores based on probability distributions across ratings. The example prompt is provided in the supplementary material website.

To validate these scores, we examined the model's justifications for its ratings. For instance, when assigning a maximum imagery score to "it's the ocean on fire,"$^{\diamond}$[2] the model highlighted how the juxtaposition of opposing elements (ocean/fire, calmness/chaos) creates an evocative metaphorical image that sparks imagination through its contrast.

Surprisal. Surprisal measures word predictability as the negative log probability of a word given its context [20]: $s(w_t) = -\log_2 P(w_t|\mathbf{w}< t)$, where higher values indicate less predictable words. Using GPT-2 [16] to approximate probabilities, we compute both word-level surprisal values and sentence-level averages. For instance, "and all we want"$^{\diamond}$ has lower surprisal (more predictable) than "breathe echoing the meaning of us now"$^{\diamond}$ (higher surprisal, greater novelty).

Contextual Entropy. Contextual Entropy quantifies word unpredictability in context [20], computed as: $h(w_t) = \sum_{w_i \in V} p(w_t|\mathbf{w}< t) \log_2 P(w_t|\mathbf{w}< t)$ where V is the vocabulary. Using GPT-2, we calculate sentence-level entropy from word-level values. Examples demonstrate the range: "through the mist, a voice remains"$^{\triangle}$ ($h_{avg} = 7.43$) shows high complexity, while "this is what i want"$^{\diamond}$ ($h_{avg} = 4.76$) exhibits lower entropy, correlating with cognitive processing effort [13].

Global Entropy. Global Entropy provides a holistic measure of sentence-level unpredictability, extending Contextual Entropy to capture both word-level variability and overall semantic structure. We calculate it as the probability-weighted average of word-level entropies:

$$H(S) = \frac{1}{T} \sum_{t=1}^{T} \left(h(w_t) \cdot \frac{P(w_t|\mathbf{w}< t)}{\sum_{j=1}^{T} P(w_j|\mathbf{w}_{<j})} \right), \tag{3}$$

where $H(S)$ is the sentence's Global Entropy, $h(w_t)$ is each word's entropy given its context, and the weighting factor normalizes by relative word likelihood. Using GPT-2 for probability estimation, this metric captures both predictable and novel elements in the sentence structure.

Syntactic Complexity. Syntactic Complexity quantifies a line's structural depth through constituent phrase count, measured via shallow parsing using the `nltk` library. Examples illustrate the range: "beneath the still waters, whispers of

[2] Lines marked with $\diamond$ are generated by LSTM-VAE. Lines marked with $\triangle$ are generated by LLM.

forgotten worlds rise"$^{\triangle}$ demonstrates high complexity with multiple embedded phrases, while "all you see"$^{\Diamond}$ shows minimal syntactic structure.

Imagery. Imagery measures the vividness of sensory or visual experiences evoked by language. High-scoring lines feature physical-metaphysical fusions ("i wear the dust of stars on my skin"$^{\triangle}$), sensory transformations ("shadows breathe where silence dies"$^{\triangle}$), or elemental imagery ("it's the ocean on fire"$^{\Diamond}$). In contrast, low imagery appears in abstract statements ("this is what you want to"$^{\Diamond}$) or non-sensory thoughts ("just something i think"$^{\Diamond}$).

Abstractedness. Abstractedness evaluates how a line transcends physical reality for metaphysical or conceptual meanings. High abstractedness manifests in illogical combinations ("time slips on whispers"$^{\triangle}$), personification of abstractions ("shadows breathe"$^{\Diamond}$), or sensory-conceptual fusions ("taste of time"$^{\Diamond}$). Low abstractedness appears in physical descriptions ("snow drifts in the quiet air"$^{\triangle}$) or direct statements ("i want to run here"$^{\Diamond}$).

Valence. Valence measures emotional resonance on a negative-to-positive spectrum. Highest scores go to lines evoking wonder ("eyes like mirrors of the stars"$^{\triangle}$), vitality ("eyes that dance like candle flames"$^{\triangle}$), memory and solace ("old scars tell tales of brighter days"$^{\triangle}$), or expressing direct positive states ("just free now"$^{\Diamond}$). Low valence characterizes expressions of suffering ("grief like a nail to the skin"$^{\triangle}$), isolation ("loneliness echoes off cold walls"$^{\triangle}$), or direct negative states ("i'm broken"$^{\Diamond}$), with metaphorical expressions typically receiving more extreme scores than literal statements.

Simplicity. Simplicity assesses how straightforward a line's expression is, both syntactically and semantically. High simplicity appears in two forms: minimal statements ("with you"$^{\Diamond}$, "just the wind"$^{\Diamond}$) and clear natural imagery ("snow drifts in the quiet air"$^{\triangle}$). Low simplicity characterizes lines with layered metaphors ("love writes its lies on the bones of fear"$^{\triangle}$), paradoxical combinations ("the dark breathes where the light goes blind"$^{\triangle}$), or syntactic irregularities ("then you bite the eyes that freedom from eyes"$^{\Diamond}$).

Indeterminacy. Indeterminacy measures how much a line resists immediate or singular interpretation, instead creating a space for multiple meanings. This productive ambiguity encourages readers' active engagement in meaning-making rather than passive reception. High indeterminacy appears in lines transforming abstract into concrete ("words melt like snow on skin"$^{\triangle}$), personifying abstract concepts ("the sky wears yesterday's sorrow"$^{\triangle}$), or using suggestive grammatical irregularities ("all everything we years"$^{\Diamond}$). Low indeterminacy appears in literal descriptions ("firelight flickers against the stone"$^{\triangle}$) or direct statements ("i will never be untrue"$^{\Diamond}$).

Pretentiousness. Pretentiousness measures excessive efforts to appear profound, marked by overuse of metaphors or strained language that feels self-conscious rather than naturally evocative. High pretentiousness appears in philosophical declarations ("i am the echo of your silence"$^{\Diamond}$), complex metaphysical metaphors ("machines dream in hues of pale regret"$^{\triangle}$), and stacked poetic devices ("time slips on whispers only shadows know"$^{\triangle}$). Low pretentiousness characterizes simple projective statements ("it pulls it from a hand"$^{\Diamond}$), direct

Table 1: Correlations between key features for LSTM-VAE and LLM generated lines (Lyrics dataset)

Feature Pair	**LSTM-VAE**	**LLM**
Pretentiousness – Imagery	0.748***	0.671***
Pretentiousness – Abstractedness	0.777***	0.790***
Pretentiousness – Simplicity	-0.571***	-0.563***
Pretentiousness – Global Entropy	0.156***	0.160***
Pretentiousness – Syntax Complexity	0.047	0.073*
Simplicity – Syntax Complexity	-0.202***	-0.058
Simplicity – Indeterminacy	-0.591***	-0.471***

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

expressions ("with you"$^{\diamond}$), or effortlessly poetic descriptions ("the sun dips low on endless fields"$^{\triangle}$). Notably, dramatic philosophical expressions ("in silence, i become the void"$^{\triangle}$) may not necessarily be pretentious if they feel authentic.

8 Analysis of AI-Generated Poetic Lines

This section presents a comparative analysis of poetic lines generated by LSTM-VAE and LLM. We will look at their linguistic and stylistic characteristics from the perspective of both quantitative metrics and qualitative assessment. The analysis reveals distinct patterns in the outputs from these systems, with implications for their potential role in artistic inspiration.

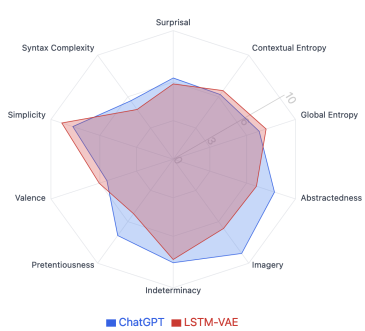


Fig. 1: Scores for LSTM-VAE poetic lines (red) and LLM lines (blue).

8.1 Quantitative Analysis

Surprisal and Entropy. LLM-generated lines show higher mean surprisal (6.318) compared to LSTM-VAE lines (5.858), suggesting that LLM tends to produce less predictable word sequences (Figure 1). However, LSTM-VAE lines exhibit higher global entropy (7.599 vs 7.033), indicating greater overall uncertainty in their language patterns. This apparent contradiction emerges from their different approaches to linguistic innovation: while LLM may choose less predictable individual words, LSTM-VAE creates more variable and unpredictable overall linguistic structures. This finding is also corroborated by the analysis of syntactic complexity: the set of LSTM-VAE lines has 685 unique syntactic patterns, however the LLM lines have only 457 unique patterns.

Lines with unusual syntactic constructions in LSTM-VAE show characteristic surprisal patterns: "clouds humanity without light" (surprisal: 6.2), "the tears above the surface trap" (surprisal: 6.4). These lines break conventional grammatical patterns while maintaining semantic suggestiveness.

In LLM's output, higher global entropy appears in lines with multiple metaphorical elements: "through the mist, a voice remains" (global entropy: 8.60), "lost in the static between breaths" (global entropy: 8.42).

Imagery and Abstractedness. A striking difference appears in the imagery scores, with LLM lines scoring significantly higher (mean 9.062) compared to LSTM-VAE lines (6.663). LLM also produces more consistently abstract language (mean 8.284) compared to LSTM-VAE (6.809). These differences reflect LLM's tendency to employ conventional poetic devices and metaphorical language, while LSTM-VAE generates more varied and potentially unconventional imagery.

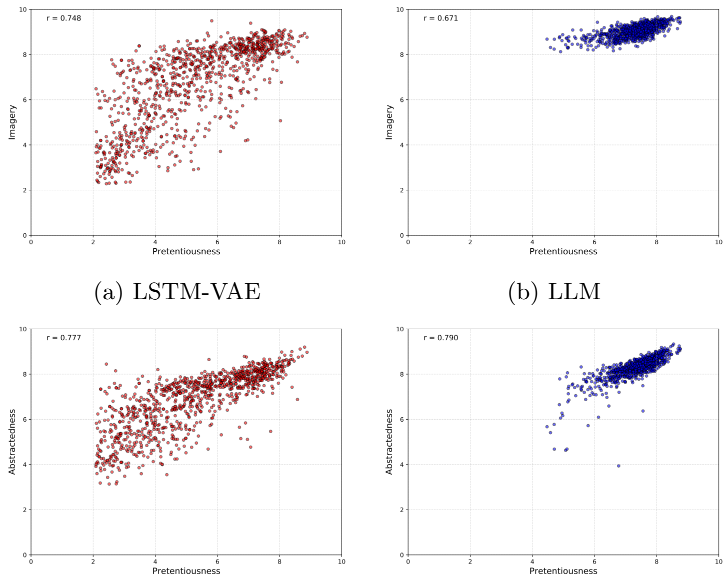
LSTM-VAE lines often blend abstract and concrete elements, leading to moderate abstractedness scores (6.8 average): "i feel through the universe" (combines physical sensation with cosmic abstraction), "through the wrong moon" (juxtaposes concrete imagery with conceptual impossibility).

In contrast, LLM maintains consistently high abstractedness: "in silence, i become the void" (abstractedness: 9.11), "somewhere, dreams rust in forgotten fields" (abstractedness: 8.99).

Pretentiousness and Complexity. LLM lines show notably higher pretentiousness scores (7.339) compared to LSTM-VAE lines (5.245). This correlates strongly with abstractedness ($r=0.790$ for LLM, $r=0.777$ for LSTM-VAE) and imagery ($r=0.671$ for LLM, $r=0.748$ for LSTM-VAE), as shown in Table 1 and the scatterplots in Figure 2. Even when dealing with traditionally "poetic" themes, LSTM-VAE lines maintain relatively low pretentiousness scores: "we are the edge" (pretentiousness: 5.3), "the moon is bare" (pretentiousness: 4.8).

LLM's lines show a clear correlation between imagery and pretentiousness: "love writes its lies on the bones of fear" (imagery: 9.56, pretentiousness: 8.27), "moonlight drips from broken clouds" (imagery: 9.23, pretentiousness: 7.53).

Simplicity and Indeterminacy. A notable pattern in LSTM-VAE output is the balance between simplicity and indeterminacy. Lines like "i feel it all" and



(a) LSTM-VAE (b) LLM

(c) LSTM-VE (d) LLM

Fig. 2: Comparison of Pretentiousness vs. Imagery (a, b) and Pretentiousness vs. Abstractedness (c, d) correlations in LSTM-VAE (red) and LLM (blue) generated lines.

"beneath the moon" achieve high simplicity scores (>9.0) while maintaining moderate indeterminacy (7.8), suggesting that syntactic simplicity does not preclude interpretive openness.

Among the lines output by LLM, complex metaphorical lines have an inverse relationship between pretentiousness and simplicity: "through empty streets, we wander alone" shows high simplicity (9.0) but lower pretentiousness (6.61), "love writes its lies on the bones of fear" shows high pretentiousness (8.27) but low simplicity (4.91).

8.2 Qualitative Analysis

LSTM-VAE Generated Lines. LSTM-VAE lines exhibit several distinctive characteristics that contribute to their potential as creative stimuli:

1. **Syntactic Indeterminacy**: Lines often feature incomplete or unconventional syntactic structures: "we're lost and happiness in the way", "beneath the longing". These fragmentary constructions create spaces for interpretation and completion.
2. **Semantic Openness**: Many lines balance concrete imagery with abstract concepts: "i feel the universe of the world steel", "through the dream and

can not see". Such combinations of tangible and intangible elements invite multiple interpretations.

3. **Novel Combinations**: The system creates unexpected juxtapositions that often maintain semantic suggestiveness, but break habitual patterns: "clouds humanity without light", "the tears above the surface trap".

LLM Generated Lines. LLM's output demonstrates different characteristics:

1. **Conventional Poetic Devices**: Lines frequently employ traditional poetic techniques: "Memory bleeds like ink across a broken page", "The moon is a scar in the velvet sky". These constructions, while evocative, often rely on familiar metaphorical patterns.
2. **Syntactic Completeness**: Lines tend to be grammatically complete: "The night holds me close, but won't let me stay", "Through empty streets, we wander alone". This completeness may limit their potential for creative reinterpretation.
3. **Thematic Coherence**: Lines maintain consistent imagery and themes: "The silence burns with unspoken truths", "We are fragments of forgotten songs". While aesthetically pleasing, this coherence may constrain creative exploration.

9 Implications for Creative Inspiration

Our analysis reveals fundamental differences in how these systems generate text that might serve as creative stimuli. These differences manifest across several key dimensions of poetic expression, particularly in their approaches to metaphor, complexity, and emotional content.

In terms of metaphorical expression and complexity, LSTM-VAE and LLM take notably different paths. LSTM-VAE gravitates toward simple, direct imagery while building complexity through syntactic disruption and semantic openness. This creates productive tensions that invite exploration and reinterpretation. In contrast, LLM constructs elaborate, multi-layered metaphors using conventional poetic devices, resulting in more traditionally "poetic" but potentially less generative outputs with predetermined meanings.

Perhaps most telling is how the two systems approach emotional expression. LSTM-VAE frequently conveys emotion through simple, direct statements like 'i feel it all', avoiding ornate constructions in favor of raw, unembellished expression. Yet this apparent simplicity often harbors a deeper semantic openness. Lines like 'beneath the longing' or 'through the dream' may appear straightforward, but their projective nature invites artists to form their own interpretations and meanings. In contrast, LLM tends to embed emotional content in elaborate metaphorical constructions that can verge on the pretentious, as evidenced in lines like 'love writes its lies on the bones of fear' (pretentiousness score: 8.27). Such overwrought metaphors create the impression of "trying too hard" and feel contrived rather than emotionally genuine.

LSTM-VAE's higher global entropy, combined with its tendency toward syntactic indeterminacy and novel combinations, suggests it may be more effective at generating lines that occupy the productive middle ground between randomness and predictability associated with the U-shaped effect hypothesis. LLM's output, while rich in poetic imagery, tends toward conventional forms and complete syntactic structures, making its lines more immediately appealing but potentially less effective as creative seeds. These patterns point to fundamental architectural differences in how these systems process and generate language, with important implications for their role in the creative process.

10 Conclusion

The relationship between computational systems and artistic inspiration requires a deeper understanding of how different approaches to text generation can either enhance or constrain creative thinking. Our comparative analysis of LSTM-VAE and LLM reveals insights that extend beyond technical performance to fundamental questions about the nature of creative stimuli.

The distinctive characteristics we observed—LSTM-VAE's tendency toward open-ended expressions versus LLM's more conventionally structured outputs —align with historical artistic approaches like the cut-up technique and theoretical frameworks such as Shklovsky's concept of defamiliarization. These connections suggest that effective creative stimuli often emerge from disrupted rather than conventional linguistic patterns.

Our results suggest that the architecture and training approach of generative models significantly influence their potential as tools for inspiration. Smaller, more focused models trained on curated datasets might be more effective at generating evocative and projective lines than larger, general-purpose language models. This is particularly relevant during the Seed phase of creativity, where the goal is not to produce polished artistic output but to help artists enter a state of heightened perception and creative flow.

Future work could build on our findings in several ways: developing more granular metrics for measuring semantic openness and emotional authenticity in generated text, conducting user studies to understand how different types of generated content influence artists' creative processes, and investigating how architectural choices in AI systems affect their ability to generate productively indeterminate content.

Beyond poetry and lyrics, our findings have implications for how we approach the challenge of computational creativity support. Our analysis reveals a fundamental contrast between two approaches to text generation: one favoring conventional coherence and fully formed meanings, the other embracing semantic-syntactic disruption and indeterminacy. This distinction suggests an alternative direction to the current trend of increasingly large and general models. Instead, we might focus on developing specialized architectures that preserve the productive ambiguity and semantic openness that appear crucial for stimulating human creativity.

References

1. Althuizen, N.: Revisiting Berlyne's inverted U-shape relationship between complexity and liking: The role of effort, arousal, and status in the appreciation of product design aesthetics. Psychology & Marketing **38**(3), 481–503 (2021). https://doi.org/https://doi.org/10.1002/mar.21449, https://onlinelibrary.wiley.com/doi/abs/10.1002/mar.21449
2. BBC: Cracked actor (excerpt). YouTube video (1975), https://youtu.be/6nlW4EbxTD8?si=4LHeAmla2DTzN8xH, accessed: Jan 23, 2025
3. Berlyne, D.E.: Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation. Journal of Aesthetics and Art Criticism **34**(1), 86–87 (1975). https://doi.org/10.2307/428656
4. Burroughs, W.: The Job: Interviews with William S. Burroughs. Penguin Modern Classics, Penguin Books Limited (2012), https://books.google.ca/books?id=DivGjW-qEv8C
5. Cross, C.R.: Heavier Than Heaven: A Biography of Kurt Cobain. Hyperion, New York (2001)
6. Gold, B.P., Pearce, M.T., Mas-Herrero, E., Dagher, A., Zatorre, R.J.: Predictability and uncertainty in the pleasure of music: A reward for learning? Journal of Neuroscience **39**(47), 9397–9409 (2019). https://doi.org/10.1523/JNEUROSCI.0428-19.2019, https://www.jneurosci.org/content/39/47/9397
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (Nov 1997). https://doi.org/10.1162/neco.1997.9.8.1735, https://doi.org/10.1162/neco.1997.9.8.1735
8. Holzer, J.: Truisms. Photostat, The Museum of Modern Art, New York (1978), https://www.moma.org/collection/works/63755, accessed: Jan 23, 2025
9. Jones, D.: David Bowie: A Life. Crown Archetype, New York, first edition edn. (2017)
10. Kingma, D.P., Welling, M.: Auto-Encoding Variational Bayes. In: 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings (2014)
11. Klütsch, C.: Computer graphic—aesthetic experiments between two cultures. Leonardo **40**(5), 421–425 (10 2007). https://doi.org/10.1162/leon.2007.40.5.421, https://doi.org/10.1162/leon.2007.40.5.421
12. Kruger, B.: I shop therefore i am. Photolithograph on paper shopping bag, The Museum of Modern Art, New York (1990), https://www.moma.org/collection/works/64897, accessed: Jan 23, 2025
13. Lowder, M.W., Choi, W., Ferreira, F., Henderson, J.M.: Lexical predictability during natural reading: Effects of surprisal and entropy reduction. Cognitive Science **42**(S4), 1166–1183 (2018). https://doi.org/https://doi.org/10.1111/cogs.12597, https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.12597
14. Molnar, V.: Molnar on generative art. Vimeo video (2019), https://vimeo.com/372579247, accessed: Jan 23, 2025
15. Pepperell, R.: Art, perception and indeterminacy. Contemporary Aesthetics **5** (2007), http://hdl.handle.net/2027/spo.7523862.0005.008
16. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al.: Language models are unsupervised multitask learners. OpenAI blog **1**(8), 9 (2019)
17. Rubin, R.: The creative act: a way of being. Penguin (2023)
18. Sahu, G., Vechtomova, O.: Computational modeling of artistic inspiration: A framework for predicting aesthetic preferences in lyrical lines using linguistic and stylistic features (2024), https://arxiv.org/abs/2410.02881

19. Schwitters, K.: Myself and My Aims: Writings on Art and Criticism. University of Chicago Press, Chicago (2021), https://press.uchicago.edu/ucp/books/book/chicago/M/bo17998666.html
20. Shannon, C.E.: A mathematical theory of communication. The Bell system technical journal **27**(3), 379–423 (1948)
21. Shklovsky, V.: Art as technique. In: Lemon, L., Reis, M. (eds.) Russian Formalist Criticism. University of Nebraska Press, Nebraska (1965)
22. Turner, C.: The deliberate accident in art (2011), https://www.tate.org.uk/tate-etc/issue-21-spring-2011/deliberate-accident-art, accessed: 2024-11-15
23. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems. vol. 30, pp. 5998–6008 (2017)
24. Vechtomova, O., Bahuleyan, H., Ghabussi, A., John, V.: Generating lyrics with variational autoencoder and multi-modal artist embeddings (2018), https://arxiv.org/abs/1812.08318
25. Vechtomova, O., Sahu, G., Kumar, D.: Lyricjam: A system for generating lyrics for live instrumental music. In: Proceedings of the 12th Conference on Computational Creativity (2021)
26. Zeilinger, M.: The Politics of Visual Indeterminacy in Abstract AI Art. Leonardo **56**(1), 76–80 (02 2023). https://doi.org/10.1162/leon_a_02291, https://doi.org/10.1162/leon_a_02291