

Is It Happening? Lyotard’s Sublime and the Indeterminate in AI-Generated Text

Olga Vechtomova

University of Waterloo, Canada
ovechtom@uwaterloo.ca

Abstract

How do we distinguish between AI-generated text that performs the appearance of creative novelty and text in which something genuinely unforeseen has arisen? We propose a framework grounded in Jean-François Lyotard’s aesthetics of the sublime, which distinguishes between art that bears witness to the indeterminate (what appears when a determining system encounters its own limit) and art that merely recombines established formulae to simulate novelty. We map this distinction onto three modes of AI text generation and evaluate them using an LLM judge that assesses, blind to source, whether texts feel crafted toward an effect or arrived at their qualities without design. Standard LLM generation embodies Lyotard’s institutional logic: its outputs are perceived as almost maximally crafted and categorizable. LSTM-VAE generation occupies a position of constitutive disarmament: the judge perceives its outputs as unanticipated, using vocabulary of accident and emergence never applied to LLM outputs. Activation steering creates fissures within the LLM’s determination: the judge perceives steered text as “stumbling into” effects rather than designing them, producing categorical friction, recognizable as prose yet containing elements no established mode accounts for. The event itself (the moment when habitual understanding stalls and something opens) is always human: it occurs when what arises within the system’s gaps meets a perceiver whose determinations are suspended.

Introduction

When an AI system generates a poem, a story, or a line of lyrics, how do we know whether something genuinely unexpected has occurred or whether the system has simply produced a well-crafted approximation of what unexpected text is supposed to look like? Current evaluation frameworks in computational creativity assess properties of the output: novelty, coherence, surprise, quality. These are useful measures, but they cannot distinguish between a system that performs creative effect and a system in which something arises that the system itself could not have anticipated. A line that scores high on novelty may still be entirely predictable from the system’s perspective. It is the most probable way to be improbable, given the training data.

This gap in our evaluative vocabulary points to a deeper question: not what the output looks like, but what relation-

ship the system bears to its own production. Does the system determine its output so thoroughly that nothing escapes its learned patterns? Or are there gaps (points of incompleteness or internal conflict) where something genuinely undetermined can emerge?

The French philosopher Jean-François Lyotard addressed a structurally similar question in the domain of art. In “The Sublime and the Avant-Garde” (Lyotard 1991), he distinguished between art that bears witness to what he called the indeterminate (something that escapes all rules, categories, and institutional predictions about what should come next) and art that merely innovates, recombining “formulae confirmed by previous success” to produce novelty without ever confronting what lies outside its determining logic. For Lyotard, the indeterminate is not a property that certain artworks possess. It is what appears when a determining system (a tradition, an institution, a set of rules) encounters its own limit. The encounter with it is what he calls the sublime: a pre-cognitive feeling in the perceiver at the moment when habitual understanding stalls and something opens.

Lyotard developed these ideas partly through visual art, but his framework is grounded in the logic of linguistic succession (what sentence follows what sentence, what rules govern the linking) and is therefore directly applicable to text generation. We argue that his framework maps onto the landscape of AI generative systems with unexpected precision, distinguishing three modes of generation according to where and how the indeterminate arises (or is prevented from arising) within the generative process. We draw on empirical findings from two studies: a comparative analysis of LSTM-VAE and LLM outputs (Vechtomova and Sahu 2025), and an investigation of activation steering for creative estrangement (Vechtomova et al. 2026).

The Sublime, Disarmament, and the Institution

We now outline the key elements of Lyotard’s framework before applying them to generative architectures in Section 3. Lyotard’s aesthetics of the sublime centers on a question that precedes all others: not “What is happening?” but “Is it happening?” which is the sheer fact of occurrence before any determination of content or meaning. “What we do not manage to formulate is that something happens. . . Or rather,

and more simply, that it happens. . . Not a major event in the media sense, not even a small event. Just an occurrence” (Lyotard 1991). The event is “infinitely simple, but this simplicity can only be approached through a state of privation”, a stripping away of what we rely on to make sense of experience. To encounter it, “that which we call thought must be disarmed”, freed of its programmes, its anticipatory frameworks, its institutional presumptions about what should come next.

The indeterminate in Lyotard’s account is not a quality that certain artworks possess. It is the condition of the event itself. Before one can ask what is happening, what it means, whether it is good or bad, something must first occur. That occurrence, in the instant before it is named or categorized, is pure indeterminacy, the moment before the mind catches up and supplies an interpretation. “Is it happening?” is a question that cannot be answered without destroying what it asks about, because to answer is to determine, and to determine is to move past the event into interpretation. “With the advent of the aesthetics of the sublime, the stake of art in the nineteenth and twentieth centuries was to be the witness to the fact that there is indeterminacy” (Lyotard 1991). Not to represent indeterminate things, but to testify through the work’s own existence that determination is not total.

This indeterminacy stands in opposition to what Lyotard identifies as the logic of institutions. In a passage that will prove central to our argument, he writes: “After a sentence, after a colour, comes another sentence, another colour. One doesn’t know which, but one thinks one knows if one relies on the rules that permit one sentence to link up with another, one colour with another, rules preserved in precisely those institutions of the past and future. . . The School, the programme, the project — all proclaim that after this sentence comes that sentence, or at least that one kind of sentence is mandatory, that one kind of sentence is permitted, while another is forbidden” (Lyotard 1991). Both the academy and the avant-garde, Lyotard observes, share this relationship to time: both presume to know what follows. Both “forget the possibility of nothing happening, of words, colours, forms or sounds not coming; of this sentence being the last.” The institution exists to foreclose the indeterminate, to ensure that what comes next is anticipated and therefore manageable. As we will argue, the language model represents the most thoroughgoing version of this presumption ever constructed.

Lyotard draws on Burke to characterize the affective structure of this encounter. For Burke, the sublime arises from the threat of privation, the threat that something essential will be taken away: light, language, companionship, life itself. The deepest privation is the threat that nothing further will happen at all, that occurrence itself will cease. The sublime feeling arises when this threat is held at a distance: not experienced directly (which would be pure terror) but suspended, “kept at bay, held back” (Lyotard 1991). This creates an oscillation between the familiar and its withdrawal. Crucially, the sublime is not constant strangeness. Constant strangeness is quickly domesticated, as the mind categorizes the mode and habitual processing resumes. It is the irruption of the unfamiliar within a familiar frame. The indeterminate

registers as such only against a ground of the determined, the condition Shklovsky (1965) identified as the mechanism by which art restores perception, and which Pepperell (2006) linked to heightened attention when habitual recognition is suspended.

Against the genuine occurrence, Lyotard sets innovation, the market logic that simulates the indeterminate through calculated novelty. Innovation “re-uses formulae confirmed by previous success, throws them off-balance by combining them with other, in principle incompatible, formulae, by amalgamations, quotations, ornamentations, pastiche” (Lyotard 1991). The result is the *petit frisson*. Innovation performs the indeterminate without ever confronting it: “innovating means to behave as though lots of things happened, and to make them happen. Through innovation, the will affirms its hegemony over time” (Lyotard 1991). The genuine occurrence, by contrast, defeats the will.

Lyotard does not define the indeterminate as a bounded concept. To do so would be to determine it. He characterizes it relationally: it is what exists at the limit of any system’s capacity to account for what occurs. This is what Burdman (2024) calls ‘a negative principle — the absence of a principle’ that can ‘only be experienced through feeling,’ and what Mafe (2009) describes as ‘profound and positive unknowing.’

These distinctions (between disarmament and institutional presumption, between suspense and constant disruption, between stripping away and elaboration) provide the framework for analyzing three modes of AI text generation that bear fundamentally different relationships to the conditions under which the indeterminate may arise.

Three Modes of Generation

We now map Lyotard’s framework onto three modes of AI text generation. The organizing question is not whether these systems produce events (the event is always human) but where and how each architecture creates or forecloses conditions for the indeterminate to arise.

To ground the framework empirically, we conducted an evaluation using Claude Opus 4.6 as judge. Text snippets from each generative condition were presented individually, without source labels, and assessed on two questions: (Q1) does the text feel deliberately crafted toward a literary effect, or do its qualities feel like they arose without deliberate design? (Q2) does the text belong to a recognizable mode of writing, or does it resist categorization? Both were scored 1–5 (Q1: 1=clearly crafted, 5=entirely unanticipated; Q2: 1=immediately categorizable, 5=resists categorization), with a forced reasoning statement before each score. Each snippet was evaluated five times at temperature 0.7, scores averaged. Two comparisons were conducted: Comparison A evaluated sentence-level outputs (100 randomly sampled lines per condition from standard LLM and LSTM-VAE pools), and Comparison B evaluated passage-level outputs (first 150 words from 12 stories generated with Llama 3.1 8B under each of the two conditions: explicit prompting for experimental prose and dual-layer activation steering).

We use an LLM as our evaluative apparatus while arguing that the LLM is the most complete realization of institutional logic. This is not a contradiction. Art critics are institutional; that does not disqualify them from registering what escapes the institution. What they cannot do is produce the work that escapes it. We use the judge accordingly: not as an oracle of the indeterminate, but as an institutional apparatus whose vocabulary shifts when it encounters outputs it cannot fully assimilate. The evidence we report is how the vocabulary shifts: from forensic confidence when judging LLM lines to genealogical conjecture when judging LSTM-VAE lines ¹.

The LLM as Institution

The Large Language Model, trained on the entire written archive of the internet, is the most complete realization of Lyotard's institutional logic ever constructed. Its training objective (predicting the next token given all preceding tokens) is the internalization of "the rules that permit one sentence to link up with another." Its determining capacity is so thorough that nearly every gap in which indeterminacy might arise is closed. When ambiguity appears, it is the kind the institution recognizes and permits: managed, conventional, itself a product of determination.

The evaluation confirms this. Standard LLM poetic lines (generated by ChatGPT-4o) scored 1.030 ± 0.193 on Q1 and 1.016 ± 0.126 on Q2. In 487 of 500 evaluations, the judge assigned a score of 1 on Q1. The reasoning texts reveal why: the judge consistently identifies an author behind the text, using vocabulary of intention and design: "deliberately composed," "carefully paired imagery," "conscious literary craft," "purposeful literary gesture." Terms such as "accidental," "unintended," or "emerged" appear in zero LLM evaluations. The judge operates forensically: it identifies literary devices (personification, metaphor, alliteration), treats them as proof of intentional design, and uses the same devices to name the tradition the text belongs to.

This institutional completeness extends to the conventions of unconventionality. When explicitly prompted to produce fragmented, experimental prose (Vechtomova et al. 2026), the LLM produces the established conventions of fragmentation (ellipses, sentence fragments, lowercase) devices that are immediately categorizable. The institution has absorbed its own transgression.

The LSTM-VAE as Disarmament

The LSTM-VAE occupies a fundamentally different position. Trained on a small curated dataset (~20,000 lyric lines) with several hundred thousand parameters, it generates not by autoregressive sequence prediction but by sampling from a continuous latent space in which nearby points encode semantic proximity rather than sequential likelihood (Vechtomova, Sahu, and Kumar 2021). The indeterminate arises within this system because the system's determining capacity is insufficient to eliminate it. The decoder cannot fully resolve interstitial latent points into conventional form. The

indeterminate is generated by the system's own incompleteness.

The evaluation captures this directly. LSTM-VAE lines scored 3.806 ± 1.183 on Q1 (vs. 1.03 for the LLM) and 2.398 ± 1.085 on Q2 (vs. 1.016), a gap of 2.78 on the crafted-vs.-arrived scale. The reasoning texts are where the distinction becomes most visible. Where the judge reasons forensically about LLM text (identifying devices as proof of craft), it reasons genealogically about LSTM-VAE text, asking what *process* produced it. "Accidental" appears in 195 of 500 LSTM-VAE evaluations and zero LLM evaluations; "emerged" in 83 vs. zero. The judge reaches for origin-hypotheses (a mishearing, a translation error, an overheard fragment, automatic writing) rather than identifying devices.

The contrast is sharpest when comparable texts receive opposite judgments. The LLM line "emptiness wears a velvet cloak" is scored Q1=1, Q2=1: "a concise, deliberate metaphor, clearly crafted." The LSTM-VAE line "the world's burns" is scored Q1=5, Q2=5: "likely a grammatical error... accidental ambiguity... entirely unintended." Both achieve poetic effect through unusual language. The judge applies device-forensics to one and origin-diagnosis to the other. Similarly, for "a longing of heaven," the judge observes that the preposition "of" instead of "for" accidentally makes heaven the one longing. In each case, the judge describes exactly what our framework predicts: the indeterminate arising in the gap of the system's incompleteness, producing qualities no one aimed at.

Not all LSTM-VAE outputs achieve this quality. Many are unremarkable, some incoherent. This inconsistency is essential. If the system reliably produced the indeterminate, its incompleteness would have become a programme. The LSTM-VAE satisfies Lyotard's condition through what he calls privation, the stripping away of capacity itself. It operates, in his phrase, *ex minimis*.

Activation Steering: Fissures in the Determined

The third mode asks whether the conditions for the indeterminate can be deliberately created within a system that has already internalized the institution.

Activation steering modifies a language model's internal representations during generation by adding directional vectors to hidden states (Turner et al. 2023). In Vechtomova et al. (2026), steering vectors constructed from contrastive pairs embodying the aesthetics of Tarkovsky and Lynch were applied to Llama 3.1 8B. A dual-layer approach applied a compositional vector at one layer and a content-level vector at an adjacent layer, separating how meaning is constructed from what reality is treated as stable. The steered outputs exhibit emergent qualities absent from all training data (temporal self-contradiction, identity instability, ontological uncertainty):

"Emily stood on the edge of the cliff, staring at the cliff. She had been staring at the cliff for hours. They said the cliff was not there."

The indeterminate has arisen because two determining systems (linguistic competence and ontological consistency) have been forced into misalignment. The grammar permits

¹Full evaluation prompts, reasoning texts, line samples, and score distributions available at: <https://ovechtom.github.io/lyotard-supplementary/>

the sentence; the world-model should forbid it. In the fissure, something arises that neither would have produced alone. This is not disruption of syntax but disruption through syntax (Vechtomova et al. 2026). The Emily passage above shows what this means concretely: short sentences, plain vocabulary, no ornament. The indeterminate is emerging here through what Lyotard calls “extreme simplicity of turn of phrase” (Lyotard 1991) rather than elaboration.

Comparison B evaluated steered passages (dual-layer, L14 $\alpha=4.5$) against prompted experimental prose (three prompting modes: experimental, stream-of-consciousness, fragmented). On Q1, steered outputs scored 2.60 vs. 1.53 for prompted. The reasoning texts reveal the nature of this gap. The judge describes prompted outputs as “a deliberate exercise” (16 of 60 evaluations; zero for steered), “consciously crafting a literary effect,” and “deliberately crafted to simulate a particular mental state.” For steered outputs, the vocabulary shifts: “stumbles into” appears in 7 evaluations (zero for prompted), “accidental” in 12 vs. zero, “losing coherence” in 5 vs. zero. The prompted text is recognized as a competent performance of a mode. The steered text is recognized as something that does not fully control what it is doing. It is through that loss of control that the indeterminate arises.

On Q2, the two conditions scored nearly identically (1.97 vs. 1.90), but the reasoning texts reveal a qualitative distinction the scores compress. The prompted text is categorized with confidence: “stream-of-consciousness” appears in 54 of 60 prompted evaluations vs. only 9 for steered. The judge knows exactly what mode it is looking at. For steered text, the dominant category is “literary fiction” (57/60), but it is qualified with terms that signal discomfort: “uncanny” (5 vs. 0), “contradictions” (6 vs. 0), “surreal” (21 vs. 5), “glitch” (5 vs. 0), “malfunction” (7 vs. 0). The judge categorizes the frame (literary fiction) but keeps flagging that something within the frame does not belong: “odd internal contradictions,” “not itself an established literary mode,” “feels more like a malfunction than an intentional literary mode.” The steered text creates what we might call *categorical friction*: it is recognizable enough as prose to be filed, but what it contains does not fit the file.

Activation steering thus represents what Lyotard describes as “the artist attempts combinations allowing the event.” The researcher sets conditions with precision; what arises at the points of misalignment exceeds those conditions. The indeterminate arose in the fissure. Whether it constitutes an event depends on what happens when it reaches a human perceiver.

Discussion

The three modes of generation are not a quality hierarchy. They represent different relationships to the indeterminate: a system in which determination is so thorough that indeterminacy has almost no space to arise (the LLM), a system in which the indeterminate arises in the gaps of incompleteness (the LSTM-VAE), and a system in which determination is selectively disrupted and the indeterminate arises in the resulting fissures (activation steering). The indeterminate has

no single origin. It has sites of emergence, places where the determining apparatus fails to be total.

The evaluation results suggest that these sites leave detectable traces. The LLM judge (itself an institution) consistently perceived LLM outputs as crafted and LSTM-VAE outputs as arrived at without design, using entirely different vocabularies for the two (“deliberately composed” vs. “accidental,” “emerged”). For the steered outputs, it perceived something the prompted outputs did not produce: text that “stumbles into” effects rather than designing them, and that creates categorical friction, recognizable as prose yet containing elements the judge flags as “uncanny” or “surreal.” That an LLM can detect the difference between performed and arisen indeterminacy is itself significant: it means the determining process leaves a trace in the artifact that even the institution’s own apparatus can register.

But the indeterminate arising within a system is not yet the event. The event is always human because it requires a layer beneath language, which the LLM does not have. The event occurs when indeterminate output meets a perceiver whose habitual recognition is interrupted, when someone encounters “They said the cliff was not there” and their meaning-making apparatus stalls, and in that stall something opens.

The event requires disarmament on both sides: the system’s determination must be incomplete, and the perceiver’s determination must be suspended. In artistic practice, this manifests as a chain of availability: the generative system produces something it could not foresee; the artist, attending without preconception, recognizes that something has arisen, not through analysis but through a pre-verbal response that precedes interpretation. The artist’s encounter is where intention meets its own limit: conditions are set up, processes initiated, and then something arises that was not chosen. It has a quality, a weight, a rightness, but the artist did not put it there. It appeared in the gap between the artist’s determining activity and the system’s own behaviour. The artist’s role is not to produce the event but to be available when it occurs, and to have the discipline not to override it with a plan. This is disarmament as creative practice. At each site in the chain (the system’s gap, the artist’s recognition, the listener’s encounter) the indeterminate arises anew.

We acknowledge a paradox inherent in this enterprise. Any theoretical account of the indeterminate participates in the institutional determination it describes. This is what Lyotard, invoking Orpheus and Eurydice, warned against: the turn of the head that causes what was present to withdraw. We can describe where the determining apparatus is thicker and where it has cracks. We cannot describe what arises in those cracks without foreclosing it. This account identifies architectural conditions that foreclose the indeterminate, and it can guide the design of systems that leave more gaps, but it cannot guarantee the event, because the event is what occurs when guarantee fails. Rather than asking whether AI can be creative, we might ask whether different architectures permit something to arise within them that they did not anticipate and cannot assimilate, something that may reach the human encounter in a form that has not yet been determined.

References

- Burdman, J. 2024. Lyotard and democratic aesthetics: The sublime, the avant-garde, and the unrepresentable. *Journal of Social and Political Philosophy* 3(1):37–51.
- Lyotard, J.-F. 1991. The sublime and the avant-garde. In *The Inhuman: Reflections on Time*. Stanford, CA: Stanford University Press. 89–107.
- Mafe, D. 2009. Art and the sublime: The paradox of indeterminacy, unknowing and (dis)orientation in the presentation of the unrepresentable. *Ejournalist* 9(1):58–69.
- Pepperell, R. 2006. Seeing without objects: Visual indeterminacy and art. *Leonardo* 39(5):394–400.
- Shklovsky, V. 1965. Art as technique. In Lemon, L. T., and Reis, M. J., eds., *Russian Formalist Criticism*. Lincoln, NE: University of Nebraska Press.
- Turner, A. M.; Thiergart, L.; Leech, G.; Udell, D.; Vazquez, J. J.; Mini, U.; and MacDiarmid, M. 2023. Steering language models with activation engineering. *arXiv preprint arXiv:2308.10248*.
- Vechtomova, O., and Sahu, G. 2025. Balancing indeterminacy and structure: Neural text generation for artistic inspiration. In *Proceedings of the 12th International Conference on Artificial Intelligence in Music, Sound, Art and Design (EvoMusArt)*.
- Vechtomova, O.; French, G.; Schneider, X.; and Poulin, O. 2026. Activation steering for creative estrangement: Generating evocative narrative through controlled disruption of language models. *Proceedings of the 17th International Conference on Computational Creativity*.
- Vechtomova, O.; Sahu, G.; and Kumar, D. 2021. LyricJam: A system for generating lyrics for live instrumental music. In *Proceedings of the 12th Conference on Computational Creativity*.