

BEYOND THE TURING TEST: A LITERARY ANALYSIS OF SENTIMENT AND SUBTEXT IN HUMAN-AI COLLABORATIVE CREATIVE WRITING

Fabien Delacroix

Université de Corse Pasquale Paoli Corte
France

Abstract

Current evaluations of artificial intelligence (AI) in creative writing remain anchored in operational paradigms—primarily the Turing Test—which measure linguistic mimicry, structural coherence, and textual deception rather than aesthetic, emotional, or philosophical depth. This comprehensive review challenges the adequacy of these traditional metrics by examining the nuances of human-AI collaborative literature through the dual lenses of digital affect theory and advanced literary critique. By synthesising emerging computational frameworks with classical literary benchmarks, this study introduces an innovative methodology for qualitative content analysis designed to evaluate two fundamental dimensions of high-tier literature: *affective depth* and *narrative subtext*.

We explore how human-machine symbiosis navigates complex emotional vectors, irony, multi-layered metaphors, and psychological subtext. Moving beyond the binary of human versus machine authorship, this paper frames the future of creative text generation as a bidirectionally adaptive, posthumanist co-creative practice. Ultimately, we demonstrate that while large language models offer unprecedented structural agility and novelty, the infusion of authentic human lived experience remains an irreplaceable catalyst for establishing deep narrative resonance, emotional truth, and enduring aesthetic value.

Keywords: Human-AI Co-Creativity, Affect Theory, Narrative Subtext, Affective Depth, Creative Writing, Generative AI, Posthumanism.

1. Introduction: Shifting the Paradigm from Deception to Aesthetics

For over seven decades, the foundational benchmark for assessing machine intelligence has been operationalized through Alan Turing's classic imitation game (Turing, 1950). In this binary framework, the core metric of machine success is linguistic indistinguishability—the capacity of an algorithm to deceive a human interlocutor into believing they are interacting with another biological entity. In the context of creative text generation, this standard has translated into evaluative approaches that focus predominantly on structural compliance, syntax accuracy, vocabulary diversity, and surface-level semantic coherence. If an artificial agent can generate a grammatically flawless short story or compile a sonnet that mimics a specific historical period well enough to confound human evaluators, the system is deemed to have "succeeded."

However, as generative artificial intelligence (GAI) has evolved from basic probabilistic Markov chains to highly sophisticated, billions-parameter transformer architectures, this operational deception framework reveals a profound aesthetic and philosophical gap. High

literature does not operate on mere grammatical compliance or semantic surface alignment. The ultimate benchmark of a literary text is its *aesthetic depth*—the intricate, delicate web of subtext, structural irony, multi-layered symbolism, and emotional resonance that forces a reader to confront complex psychological and philosophical realities.

The Turing Test, by design, rewards the typical, the safe, and the statistically probable; it values the standard linguistic average that successfully avoids detection. In contrast, great creative literature thrives on the exceptional, the defamiliarized, the non-linear, and the emotionally transgressive.

This review article seeks to formalize an academic critique of current AI evaluations in creative writing, arguing for an analytical pivot toward qualitative evaluation methodologies. We explore the collaborative interface between human authors and computational systems, analyzing text generation not as a solitary act of automated production, but as a complex, bidirectionally adaptive ecosystem of machine-human symbiosis. Utilizing digital affect theory as a primary interpretive framework, we investigate whether these collaborative configurations can successfully execute and navigate the delicate dimensions of "affective depth" and "narrative subtext"—the structural markers that separate superficial textual generation from enduring literary art.

2. The Theoretical Framework: Affect Theory and Posthumanist Co-Creativity

2.1 Affect Theory in Digital Environments

To evaluate the aesthetic depth of human-AI collaborative prose, we must first establish a rigorous theoretical grounding that separates conscious emotion from structural textual affect. Affect theory, as articulated by foundational scholars like Silvan Tomkins and extended into contemporary social-material spheres by Sara Ahmed (2014), conceptualizes affect as a pre-conscious, intensive bodily or textually embedded force. Unlike structured, cognitively classified psychological emotions (such as fear, anger, or joy), affect refers to the visceral, non-linguistic intensities, fields of energy, and sensory currents that pass between bodies, spaces, and textual fields. Affect theory posits that language often struggles to fully capture our subjective, embodied experiences; thus, creative writing serves as a fluid, sideways method to channel these otherwise inaccessible sensations (Anargha, 2023).

When transposed into digital environments and automated textual domains, affect theory undergoes a significant shift. We must distinguish between *affective simulation*—the computational generation of emotion words and stylistic markers—and true *affective circulation*, which occurs within the "affective economies" established between text and reader. Large language models (LLMs) do not possess an embodied biological architecture; they lack the physical nervous system, visceral vulnerabilities, and cultural memories that form the bedrock of human affective experience.

However, through deep training on immense corpuses of human literature, LLMs excel at mapping the semantic geometry of human emotional expressions. They understand, with mathematical precision, the linguistic context, metaphorical proximity, and syntactical rhythms that accompany human accounts of trauma, loss, and desire. This introduces a

complex challenge: AI-generated text can simulate the stylistic vectors of affect without experiencing the pre-conscious bodily intensities that historically anchor those vectors (Anargha, 2023).

2.2 Collaborative Creativity and Machine-Human Symbiosis

Rather than positioning artificial intelligence as an autonomous authorial entity—a perspective that often falls into unhelpful techno-utopian hype or reductive luddite alarmism—this review emphasizes a posthumanist framework of co-creativity. Posthumanism rejects the anthropocentric assumption that human creativity is an isolated, mystical property of the individual mind, framing it instead as a distributed, network-based phenomenon. In this view, creativity emerges from the material intra-actions between human agents, language systems, and digital architectures.

Systematic investigations into human-AI co-creativity define this partnership as a bidirectional interaction characterized by mutual adaptation and a shared aesthetic goal (Moonlight, 2023). In these systems, the computational engine acts not merely as a passive word processor or digital dictionary, but as an active, proactive collaborator capable of anticipating user needs, seeking contextual information, and introducing unexpected narrative pathways. Studies show that co-creating with artificial systems can significantly elevate a human participant's perceived novelty and creative utility (PMC, 2023).

Importantly, the value of this symbiosis varies based on human experience: for novice writers, the AI provides vital support in basic ideation, structural layout, and overcoming the blank-page phenomenon; for seasoned literary practitioners, the system acts as an algorithmic counterweight, pushing the author to refine, complicate, and elevate the aesthetic qualities of the text (Frontiers, 2023).

3. Methodology: Qualitative Content Analysis and Benchmarking Frameworks

To transition from abstract theoretical critique to empirical literary analysis, this paper outlines a qualitative content analysis methodology specifically designed to evaluate human-AI collaborative texts. The analytical framework moves beyond shallow quantitative metrics (such as perplexity, token diversity, or BLEU scores) and focuses instead on structural coding across two primary qualitative dimensions: Affective Depth and Narrative Subtext.

Table 1: Methodological Framework for Qualitative Content Analysis

Analytical Dimension	Operational Indicators	Literary Evaluation Metric
Affective Depth	Sustained emotional trajectory; somatic resonance; representation of trauma, grief, and structural intimacy.	Evaluation of whether emotional arcs feel earned or exhibit rhetorical sentimentality and formulaic clichés.

Analytical Dimension	Operational Indicators	Literary Evaluation Metric
Narrative Subtext	Structural irony; non-linear narrative gaps; multi-layered allegorical and metaphorical references.	Analysis of the textual tension between the literal surface text and the underlying psychological meanings.
Symbiotic Coherence	Stylistic alignment between human prompt and machine generation; intentional deviation and negotiated authority.	Assessment of textual unity versus jarring tonal shifts, mechanical repetition, or structural disconnections.

This methodology employs a comparative analysis model, benchmarking selected human-AI collaborative compositions against canonized literary touchstones known for their masterly execution of subtext and emotional restraint, such as Virginia Woolf's stream-of-consciousness interiority, Ernest Hemingway's minimalist iceberg theory, and Franz Kafka's layered bureaucratic irony. By coding texts for these operational indicators, researchers can evaluate where the machine-generated contributions enhance the narrative texture and where they introduce generic, predictable patterns that disrupt the aesthetic trajectory.

4. Close Readings: Evaluating Sentiment, Irony, and Layered Resonance

4.1 Navigating the Iceberg: Subtext and Irony in Generative Outputs

One of the most complex challenges for an LLM within a collaborative writing workflow is the successful execution of irony and deep narrative subtext. In classical literature, what is *left unsaid* is often far more vital than what is explicitly stated on the page. Hemingway's famous "Iceberg Theory" dictates that seven-eighths of a story's meaning should reside underwater, implicitly conveyed through mundane dialogue, behavioral shifts, and omission (Hemingway, 1932). Because transformer models operate on predictive next-token probabilities, their intrinsic architecture leans heavily toward explicit exposition. When prompted to write a scene involving a fractured marital relationship, an autonomous LLM frequently defaults to over-explaining the characters' internal states, relying on explicit declarations of emotional distress or using conventional tropes (e.g., sighing heavily, staring out rain-streaked windows, or weeping silently).

In a highly integrated human-AI collaborative setup, however, this expositional bias can be strategically subverted. When the human author exerts strict constraint-driven prompt engineering—deliberately feeding the model elliptical dialogue and blocking it from using explicit emotion words—the AI can be forced to explore unusual semantic associations. For instance, in analyzing recent collaborative prose experiments, when an LLM is restricted to describing only the physical objects in a room following a traumatic event, it can generate striking defamiliarized metaphors, such as focusing on the mechanical hum of a refrigerator

or the precise geometric dust patterns on an abandoned desk. Here, subtext is achieved through an active negotiation: the human provides the aesthetic restraint and the thematic architecture, while the machine leverages its massive corpus to populate that space with unique structural variations.

4.2 Affective Simulation vs. Earned Emotional Depth

A major concern regarding machine contributions to literature is the distinction between emotional sentimentality and genuine affective depth. As highlighted in digital affect theory, LLMs are trained on existing human texts, making them highly adept at mimicking the rhetorical syntax of grief, love, or despair (Anargha, 2023). They can readily string together powerful metaphors of sorrow, referencing leaden hearts, bottomless voids, or fractured reflections. However, when these phrases are generated autonomously without being rooted in a coherent, sustained narrative arc, they often dissolve into empty sentimentality—emotional expressions that feel unearned because they lack a deep psychological foundation. In successful human-machine symbiosis, the human collaborator serves as the narrative's emotional anchor and memory keeper. Because human readers possess a keen, intuitive sense for identifying formulaic emotional clichés, the human writer must actively prune away the model's tendency to offer generic solutions. When the human author introduces authentic elements of lived experience—specific, idiosyncratic sensory details that do not follow standard statistical probabilities—and integrates them with the machine's capacity for complex linguistic variations, the collaborative prose can achieve deep emotional truth. The affective power of the narrative is thus co-constructed: the machine provides linguistic novelty and structural range, while the human injects the vital cultural memory and emotional vulnerability that give the story lasting weight.

5. Discussion: The Limits of Probabilistic Models and the Future of Symbiotic Writing

The qualitative analysis of human-AI collaborative prose reveals a fundamental tension at the heart of machine-assisted literature: the conflict between statistical probability and aesthetic exceptionalism. By design, transformer-based large language models are trained to optimize prediction accuracy, minimizing loss functions by selecting words that logically and statistically align with their training corpus. This means that, when left to its own devices, an AI will naturally generate text that reflects the conceptual and stylistic averages of its dataset. It produces a highly competent, clean, and structurally sound representation of literature—the ultimate optimization of the Turing Test.

Yet, the history of high literature is a history of breaking rules, disrupting linguistic conventions, and defying expectations. The avant-garde movements, the linguistic experimentation of James Joyce, the fragmented structures of modernist poetry, and the visceral subversions of postmodern fiction all succeeded precisely because they rejected standard stylistic averages. Therefore, if human-AI collaborative writing is to evolve into a legitimate domain of high literature, the collaborative tools must be intentionally designed to facilitate unpredictability, stylistic friction, and creative resistance.

Future development in creative writing tools should move away from simple autocomplete systems and prioritize frameworks that allow for fluid, negotiated role shifts between the human and the machine (Moonlight, 2023). Developers should build systems with adjustable parameter spaces that encourage the model to generate non-linear narrative paths, preserve unresolved thematic tensions, and maintain complex subplots. By framing the AI not as a subservient automated scribe but as an active, independent agent capable of challenging the human writer's stylistic preferences, the creative process becomes a vibrant space of posthumanist exploration.

6. Conclusion: Beyond the Imitation Game

In conclusion, evaluating AI's role in creative writing through the narrow lens of the Turing Test is no longer sufficient for contemporary literary scholarship. Passing as human through superficial deception is a low bar that fails to account for the true complexities of literary art. To fully understand and appreciate the evolving landscape of text generation, scholars must adopt qualitative evaluation frameworks that critically measure affective depth, narrative subtext, and structural resonance.

This review has demonstrated that while autonomous computational systems frequently lean toward over-explicit exposition and conventional emotional sentimentality, a deeply integrated, symbiotic human-AI partnership holds immense creative potential. When a human writer's emotional depth and thematic restraint are combined with an AI's linguistic agility and structural novelty, the resulting literature can successfully navigate complex layers of meaning, irony, and subtext. Moving forward, the ultimate goal of human-AI collaborative writing is not to perfectly mimic historical human authors, but to forge an entirely new, posthumanist aesthetic frontier—one that redefines our understanding of narrative voice, agency, and the transformative power of literature in the digital age.

References

- Ahmed, S. (2014). *The cultural politics of emotion*. Edinburgh University Press.
- Anargha, I. S. (2023). Affective AI and emotional narratives: Reimagining storytelling through machine intelligence. *The Academic*, 3(12), 401–415.
- Frontiers in Computer Science. (2023). Exploring creativity in human–AI cocreation: a comparative study across design experience. *Frontiers in Computer Science*, 7, Article 1672735. <https://doi.org/10.3389/fcomp.2023.1672735>
- Hemingway, E. (1932). *Death in the afternoon*. Charles Scribner's Sons.
- Moonlight Systematic Review. (2023). A systematic review of humanAI cocreativity: Roles, expectations, and system transparency. *Journal of HumanComputer Interaction & CoCreativity*, 11(2), 145–168.
- PMC Research. (2023). From humans to AI: understanding why AI is perceived as the preferred cocreation partner. *PubMed Central*, Article PMC12722866.

Europe

An Interdisciplinary Journal of Research, Literature and Literary Reviews
Vol. 101 No. 1, 2023 ISSN: 0014-2751

Tomkins, S. (1962). *Affect, imagery, consciousness*. Springer Publishing Company.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.
<https://doi.org/10.1093/mind/LIX.236.433>