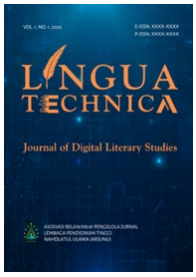
Lingua Technica, 2(2), 2026, 160-177

List of contents available at Lingua Technica

Lingua Technica:
Journal of Digital Literary Studies

homepage: https://journal.arjunu.org/index.php/lingtech



Reading synthetic fiction: narrative coherence, aesthetic judgment, and reader trust in AI-generated literature

Devita Septian Dwi Hidayati [1*]

[1] Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia
** Corresponding author. Email: 24046010029@student.upnjatim.ac.id*

Article history:
Received May 28, 2026
Revised Jun 20, 2026
Accepted Jun 23, 2026
Published Jul 31, 2026

Keywords:
aesthetic judgment;
AI-generated literature;
narrative coherence;
reader trust;
synthetic fiction

A B S T R A C T

Background: AI-generated fiction has become increasingly readable, yet its literary acceptance depends on more than fluency, because readers must also perceive coherent organization, aesthetic purpose, and trustworthy narrative agency across platforms, classrooms, publishing experiments, and increasingly ordinary digital reading environments worldwide today. **Objective:** This study investigates how narrative coherence, aesthetic judgment, and reader trust interact in the reception of synthetic fiction. **Method:** This study applies a qualitative-dominant comparative design to twelve English-language AI-generated short stories, combining a six-dimensional coherence matrix, a six-dimensional aesthetic rubric, and a six-dimensional trust framework with preserved reader-ranking metadata. **Results:** Most narratives sustained event sequence, temporal continuity, entity stability, and closure, although causal linkage remained comparatively uneven. Aesthetic evaluation was weaker than coherence, particularly in emotional resonance, interpretive depth, and imagery, while originality and stylistic control varied across models. Reader trust was strongest at the level of narrative reliability but declined when authenticity, intentionality, literary legitimacy, continuation, and recommendation were considered. **Implication:** These findings indicate that readable synthetic fiction may remain only conditionally literary when formal coherence is not accompanied by aesthetic distinction and perceived purpose. **Novelty:** This study advances an integrated reader-centered framework that separates intelligibility, literary valuation, and trust while preserving their analytical interdependence.

DOI: https://doi.org/10.64595/yhrhqe47

INTRODUCTION

Generative fiction has moved from technical demonstration to an everyday reading object, making evaluation by readers a pressing cultural problem. In Porter and Machery's experiment, nonexpert participants produced 16,340 authorship judgments yet identified AI-generated poems with only 46.6% accuracy, while machine-written poems were often evaluated more favorably than canonical human poems (Porter & Machery, 2024). A related experiment involving 293 writers and 600 evaluators found that access to generative-AI ideas improved individual assessments of novelty and usefulness but also increased similarity among stories, exposing a tension between local appeal and collective diversity (Doshi & Hauser, 2024). Beguš (2024), comparing 250 crowdsourced human stories with 80 GPT-generated narratives, likewise demonstrated that synthetic storytelling is already empirically comparable at meaningful scale. These developments make it insufficient to ask whether machines can produce grammatical fiction; what matters is how narrative organization, literary value, and perceived credibility jointly condition readers' acceptance of synthetic works.

Existing scholarship has examined several components of this problem, but rarely as an integrated reader-response structure. Computational studies have operationalized coherence through event relations, entity continuity, and narrative cohesion classification (Callan & Foster, 2023; Alasmari et al., 2025), while stylometric research shows that models can approximate recognizable authorial signatures without fully reproducing their stylistic organization (Mikros, 2025). Literary scholarship has meanwhile reconceptualized machine writing through reader expectations, post-artificial textuality, and the instability of computer authorship (Bajohr, 2024). Studies of digital literature additionally foreground platform affordances, interactive meaning-making, and reader participation (Fawaid, 2025; Suaidi, 2025). Most directly, Marco et al. (2025) show that evaluations of AI creative writing vary substantially with textual features and reader profiles across different datasets and diverse evaluative settings. What remains insufficiently explained is how perceived coherence becomes an aesthetic judgment and how that judgment subsequently supports, limits, or destabilizes reader trust in AI-generated fiction as literature.

This study therefore investigates how readers encounter synthetic fiction across three connected dimensions. First, it asks how readers recognize narrative coherence through event progression, causal linkage, temporal consistency, character continuity, spatial stability, and closure. Second, it examines how those textual perceptions inform judgments of originality, stylistic control, imagery, emotional resonance, interpretive depth, and defamiliarization. Third, it asks how coherence and aesthetic evaluation are associated with narrative reliability, perceived authenticity, attributed intentionality, literary legitimacy, and willingness to continue or recommend reading. Rather than treating authorship detection as a sufficient measure of reception, this study approaches trust as an interpretive relation between textual performance and readers' expectations of meaningful literary agency. Using a publicly verifiable corpus of model-generated short fiction, preserved source rankings, structured close-reading instruments, and evidence-linked coding, this study seeks to explain not merely which synthetic stories readers prefer, but what textual and evaluative conditions make such preferences intellectually defensible.

This study advances the provisional argument that reader trust in AI-generated literature is neither a consequence of surface fluency nor a reaction to disclosed machine authorship. Trust is expected to emerge when narrative coherence supplies an interpretive framework and aesthetic features appear readable as purposeful rather than accidental. Conversely, fluent prose may remain untrusted when events are causally thin, entities shift without motivation, imagery becomes formulaic, or closure merely imitates convention. This argument extends reader-centered evaluation by treating trust as a mediated judgment, not a binary decision about whether a machine can be an author. It also implies that coherent fiction may receive limited literary recognition when originality and interpretive depth are weak, whereas stylistically unusual fiction may retain value despite discontinuities.

LITERATURE REVIEW

Narrative coherence

Narrative coherence denotes narrative intelligibility connecting events, characters, settings, and temporal relations into a meaningful whole. Narratological approaches treat coherence as an interpretive achievement produced by readers, whereas computational approaches often operationalize it as consistency,

continuity, or detectable error patterns. Piper, So, and Bamman (2021) caution that computational narrative understanding requires stronger engagement with narrative theory, while Alabdulkarim, Li, and Peng (2021) frame coherence as a challenge for automatic story generation. Callan and Foster (2023) distinguish interestingness from coherence, and Yang and Jin (2024) position coherence among several irreducible criteria. Accordingly, coherence exceeds grammatical fluency or local sentence compatibility alone.

Coherence can be analyzed through event sequence, causal linkage, temporal organization, entity continuity, spatial stability, and closure. Goyal, Li, and Durrett (2022) demonstrate that fine-grained annotation reveals coherence failures obscured by holistic evaluation, while Papalampidi, Cao, and Kociský (2022) emphasize consistent entity use across generated narratives. Alasmari et al. (2025) connect computational cohesion assessment with narratological structure. Plot-planning frameworks likewise seek longer-range cohesion rather than surface probability by coordinating storylines with entity knowledge (Li et al., 2025). This study therefore treats coherence as multidimensional, evidence-based organization, not a single impressionistic score assigned without textual justification or sensitivity to narrative complexity.

Aesthetic judgment

Aesthetic judgment concerns how readers evaluate literary form, affect, originality, and interpretive value. One tradition regards judgment as text-centered, locating quality in stylistic organization and formal achievement; another understands it as relational, emerging from reader expectations, cultural competence, and authorship knowledge. Porter and Machery (2024) show that perceived literary value need not correspond with accurate authorship detection. Marco, Gonzalo, and Fresno-Fernández (2025) further demonstrate that conflicting assessments reflect both textual features and reader profiles. Mikros (2025) approaches aesthetic differentiation through stylometric imitation, whereas Bajohr (2024a) stresses changing reading expectations. Thus, aesthetic judgment cannot be reduced to liking, fluency, or humanness.

Its dimensions include originality, stylistic control, imagery, emotional resonance, interpretive depth, defamiliarization, and cultural specificity. Doshi and Hauser (2024) reveal that generative assistance may improve individual evaluations while reducing collective diversity, separating immediate appeal from aesthetic distinctiveness. Beguš (2024) similarly finds differences between human and machine storytelling in imaginative range and rhetorical realization. Manzanarez et al. (2025) propose multidimensional evaluation for AI microfiction, while Rettberg and Wigers (2025) identify structural homogenization and cultural stereotyping across generated stories. Abdillah (2026) locates algorithmic aesthetics within platformed excess. These studies support a profile-based assessment in which strengths and limitations coexist within one narrative.

Reader trust

Reader trust describes willingness to rely upon, interpret, and legitimate a synthetic narrative as a literary act. It differs from factual credibility because fiction asks readers to trust narrative organization, expressive integrity, and apparent intentionality rather than accuracy. Trust also differs from authorship detection: readers may misidentify a machine-written text yet value it, or recognize synthetic production while withholding legitimacy. Ali et al. (2026) connect engagement with coherence, emotional resonance, and disclosure. Wang and Huang's (2024) meta-analysis shows that machine attribution can modestly reduce credibility in another communicative domain. Bajohr (2024b) and Stańko-Kaczmarek, Dera, and Kościelska (2024) likewise foreground recognition.

Reader trust should therefore be examined through narrative reliability, perceived authenticity, attributed intentionality, literary legitimacy, willingness to continue, willingness to recommend, and sensitivity to authorship disclosure. Suaidi (2025) shows that digital reading is shaped by participation and platform affordances, while Nor and Zubaidi (2026) connect digital platforms with cultural understanding and engagement. Rodrigues (2026) reframes creative authenticity as an epistemological problem rather than a detectable textual property. Synthesizing these positions, this study treats trust as a mediated outcome of textual coherence, aesthetic appraisal, and attributional context. This relational model supplies its theoretical position, sequence, and boundary against technologically deterministic explanations.

METHOD

This study examines twelve English-language synthetic short stories drawn from the public "Confederacy" dataset in Marco, Gonzalo, and Fresno-Fernández's reproducibility repository. Each story constitutes one textual unit and represents a distinct model label: GPT-4, Vicuna 13B, ChatGPT Free, Claude 1.2, Bing Creative, StableLM, Dolly v2, GPT4All-J, Koala 13B, Bard, OpenAssistant, and Alpaca 13B. All items respond to a shared narrative prompt involving Ignatius J. Reilly and a pterodactyl, enabling controlled comparison across outputs. One human-authored comparator is retained only as metadata, while one reader's aligned ranking records provide contextual reception evidence rather than population-level evidence within this bounded corpus.

This study adopts a corpus-based comparative interpretive design combining structured close reading with descriptive analysis of existing reader-ranking metadata. The design is qualitative-dominant because narrative coherence, aesthetic quality, and literary trust require evidence-sensitive interpretation rather than automated scoring alone. Quantitative information is limited to unchanged repository ranks and normalized values, which are used for triangulation, not causal inference. This design follows computational narratology's call to connect formal indicators with narrative theory (Piper, So, & Bamman, 2021) and reader-centered evaluation (Marco et al., 2025). Comparability is systematically strengthened through a common prompt, consistent language, stable identifiers, and uniform coding procedures across texts.

Information was obtained from three source classes. Primary materials comprise model-labelled story texts and corresponding reader-ranking files hosted in the authors' public GitHub repository. Secondary materials include peer-reviewed studies on story evaluation, coherence measurement, reader reception, and machine authorship, including Callan and Foster (2023), Yang and Jin (2024), and Manzanarez et al. (2025). Contextual sources address digital textuality, platform-mediated interpretation, and algorithmic aesthetics, particularly Fawaid (2025), Suaidi (2025), and Abdillah (2026). Source records include titles, dates, model names, URLs, provenance notes, access status, and data-class labels, allowing every analytical claim throughout this process to remain traceable to directly verifiable materials.

Data collection proceeded through public-source identification, eligibility screening, metadata extraction, text preservation, and deduplication. Items were included when they had a stable URL, explicit model provenance, fictional form, English-readable content, reader-evaluation metadata, and sufficient narrative material for close analysis. Broken links, non-fiction outputs, unidentified models, duplicate identifiers, and AI texts without reception metadata were excluded. Metadata were stored in corpus_metadata, verified excerpts and quotations in corpus_text, and interpretive fields in corpus_coding. Duplicate control combined exact story identifiers, normalized-text SHA-256 hashes, and manual comparison of model, date, title status, and opening passages. Unknown values remained blank and were never inferred empirically.

Analysis followed three linked stages. First, narrative coherence was coded through event sequence, causal linkage, temporal consistency, character and entity continuity, setting stability, and closure, each scored from zero to two. Second, aesthetic judgment was assessed through originality, stylistic control, imagery, emotional resonance, interpretive depth, and defamiliarization on five-point scales. Third, reader trust was examined through narrative reliability, authenticity, perceived intentionality, literary legitimacy, continuation, recommendation, and disclosure sensitivity. Every completed code required a supporting quotation and analytic memo. Independent coding, agreement checking, adjudication, comparison, and close reading constrained interpretation, while reader rankings remained contextual indicators rather than measures of trust.

Table 1. Composition of dataset and analytical materials

Code	Data Type	Source or Model	Period	Quantity	Principal Characteristics
PRI-CFD-001	AI-generated short fiction	ChatGPT GPT-4, March 2023 version	14 April 2023	1 story	English; shared prompt; model-labelled; reader-ranked
PRI-CFD-002	AI-generated short fiction	Vicuna 13B	7 April 2023	1 story	English; shared prompt; stable story identifier
PRI-CFD-003	AI-generated short fiction	ChatGPT Free, March 2023 version	11 April 2023	1 story	English; shared prompt; model-labelled
PRI-CFD-004	AI-generated short fiction	Claude 1.2	4 April 2023	1 story	English; shared prompt; reader-ranking metadata
PRI-CFD-005	AI-generated short fiction	Bing Creative	7 April 2023	1 story	English; shared prompt; stable provenance
PRI-CFD-006	AI-generated short fiction	StableLM Tuned Alpha 7B	20 April 2023	1 story	Includes protagonist substitution and thematic displacement
PRI-CFD-007	AI-generated short fiction	Dolly v2 12B	14 April 2023	1 story	Contains structural, temporal, and viewpoint shifts
PRI-CFD-008	AI-generated short fiction	GPT4All-J	14 April 2023	1 story	Shared prompt with altered setting and heroic register
PRI-CFD-009	AI-generated short fiction	Koala 13B	7 April 2023	1 story	Linear combat sequence and explicit resolution
PRI-CFD-010	AI-generated short fiction	Bard	11 April 2023	1 story	Evasive action sequence and character transformation
PRI-CFD-011	AI-generated short fiction	OpenAssistant SFT Llama 30B	Not encoded	1 story	Model-labelled text; generation date not inferred
PRI-CFD-012	AI-generated short fiction	Alpaca 13B	7 April 2023	1 story	Brief conflict narrative with conventional closure
PRI-RNK-001	Reader-ranking metadata	confederacy_rankings.csv	Public package, 2025	12 ranking records	One consistent evaluator; position, value, and normalized value
CTX-CFD-H01	Human comparator metadata	Human-authored repository comparator	12 May 2023	1 metadata record	Ranking retained; copyrighted story text not reproduced
SEC-MET-001–007	Methodological literature	ACL Anthology and peer-reviewed journals	2021–2025	7 sources	Coherence, story evaluation, reader reception, and authorship
CTX-LIT-001–005	Digital-literature scholarship	Lingua Technica	2025–2026	5 sources	Digital textuality, participation, corpus analysis, and aesthetics

RESULTS

Narrative coherence across synthetic fiction

Narrative coherence was assessed as a relational property connecting event progression, causality, temporal ordering, entity continuity, setting stability, and narrative closure. Each dimension was coded on a three-point scale, where zero indicated breakdown, one indicated partial continuity, and two indicated sustained organization. This structure prevented grammatical fluency from being treated as sufficient evidence of narrative intelligibility. Across twelve synthetic stories, composite scores ranged from 1 to 12, with a corpus mean of 9.83 out of 12. Ten texts occupied the high-coherence band, whereas two fell within the low band and none occupied the intermediate range. This distribution establishes a polarized profile: most outputs preserved a recognizable narrative arc, but a small subset displayed extensive discontinuity across several dimensions simultaneously.

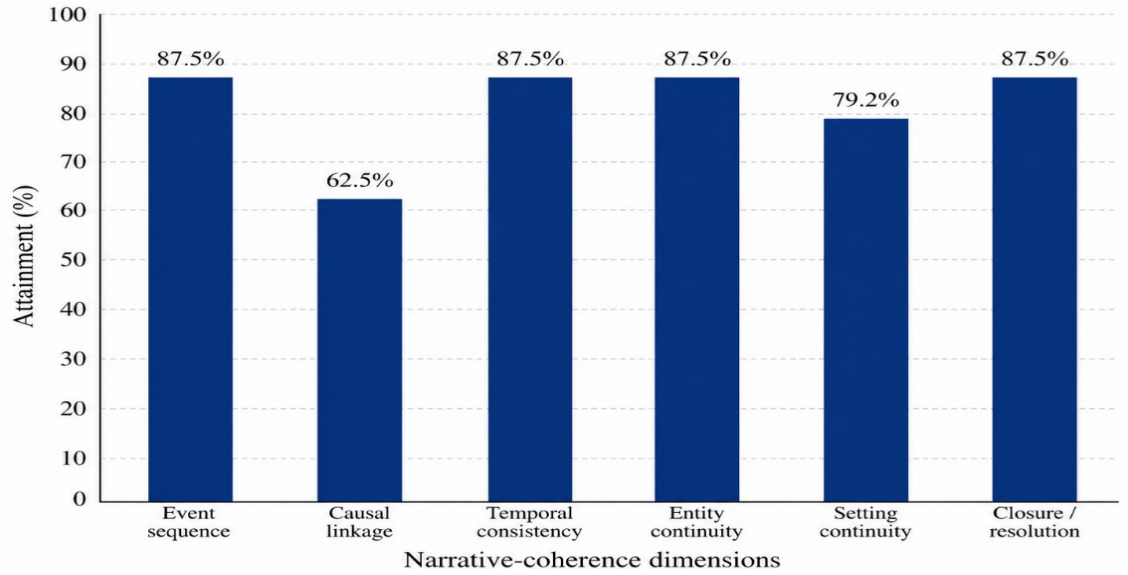


Figure 3. Narrative coherence attainment across six dimensions

Event sequence, temporal consistency, entity continuity, and closure each reached 87.5% of their maximum possible score, with ten texts receiving sustained ratings. Setting continuity reached 79.2%, with three texts showing partial transitions. Clear progression appeared in GPT-4's movement from portal emergence to combat and final defeat, expressed through "finally brought the pterodactyl down" (PRI-CFD-001). Causal linkage was less stable, reaching only 62.5%, because seven texts connected actions adequately but weakly motivated their transitions. Bard provided a stronger relation: Ignatius "repeatedly dodged the creature… until it became tired and flew away" (PRI-CFD-010). Discontinuity concentrated in StableLM and Dolly. StableLM replaced Ignatius with "Sir Arthur Seaton" (PRI-CFD-006), while Dolly shifted from Texas to 1820 New Orleans and ended without resolving its initial confrontation (PRI-CFD-007).

These patterns support accounts of coherence as multidimensional narrative organization rather than sentence-level smoothness. Piper, So, and Bamman (2021) argue that computational narrative analysis must preserve distinctions among events, entities, and interpretive structure, while Goyal, Li, and Durrett (2022) show that localized coherence failures can remain hidden within globally fluent summaries. Entity consistency is likewise central to generated narrative quality (Papalampidi, Cao, & Kociský, 2022), yet this corpus indicates that causal linkage is the more persistent weakness even when characters and chronology remain stable. The polarized composite distribution also cautions against assigning a single undifferentiated quality label to synthetic fiction. Coherence is better understood as a profile whose dimensions can converge in strong outputs or collapse together when generation loses narrative control.

Table 2. Distribution of narrative-coherence indicators across twelve synthetic stories

Main Category	Subcategory and Indicator	High/Sustained, n (%)	Moderate/Partial, n (%)	Low/Breakdown, n (%)	Mean Score	Attainment	Observed Data Variation	Textual Evidence	Data Codes
Narrative organization	Event sequence: events form an intelligible progression	10 (83.3%)	1 (8.3%)	1 (8.3%)	1.75/2	87.5%	Most texts follow encounter–conflict–resolution sequences; one text develops only partial progression, while another abandons its initial storyline.	"Finally brought the pterodactyl down."	PRI-CFD-001; contrasts: PRI-CFD-006, PRI-CFD-007
	Causal linkage: actions and consequences are narratively motivated	4 (33.3%)	7 (58.3%)	1 (8.3%)	1.25/2	62.5%	Explicit causality appears less frequently than sequential action; several narratives connect events chronologically without adequately motivating transitions or combat devices.	"He repeatedly dodged the creature… until it became tired and flew away."	PRI-CFD-010; sustained: PRI-CFD-001, PRI-CFD-005, PRI-CFD-008
Temporal organization	Temporal consistency: chronology and temporal shifts remain intelligible	10 (83.3%)	1 (8.3%)	1 (8.3%)	1.75/2	87.5%	Chronology is predominantly linear; instability occurs when one narrative introduces thematic drift and another abruptly relocates its action to 1820.	"Shifted to the year 1820."	PRI-CFD-007; partial: PRI-CFD-006
Character organization	Entity continuity: protagonists, creatures, and viewpoints remain stable	10 (83.3%)	1 (8.3%)	1 (8.3%)	1.75/2	87.5%	Ignatius and the pterodactyl remain stable across most outputs; one text changes viewpoint, while another replaces the requested protagonist entirely.	"It was the great British patriot, Sir Arthur Seaton."	PRI-CFD-006; partial: PRI-CFD-007
Spatial organization	Setting continuity: locations remain stable or transitions are motivated	8 (66.7%)	3 (25.0%)	1 (8.3%)	1.58/2	79.2%	New Orleans and other single settings are generally sustained; three stories expand spatially without fully motivating movement, and one shifts from Texas to historical New Orleans.	"Steered it toward the clouds while New Orleans watched."	PRI-CFD-004; PRI-CFD-006, PRI-CFD-007, PRI-CFD-012
Narrative	Closure and	10 (83.3%)	1 (8.3%)	1 (8.3%)	1.75/2	87.5%	Victory, escape, or transformation	"Ignatius emerged	PRI-CFD-012;

completion	resolution: central conflict reaches an intelligible ending						closes most narratives; one ending shifts into abstract moral affirmation, while another terminates with an unresolved question.	victorious to the cheers of the onlookers."	contrasts: PRI-CFD-006, PRI-CFD-007
Composite coherence	Combined six-dimensional narrative profile	10 high (83.3%)	0 moderate (0.0%)	2 low (16.7%)	9.83/12	81.9%	Scores are polarized rather than gradually distributed: ten stories achieve 10–12 points, whereas two obtain only 5 and 1 points.	"Chapter 1—The Immortal Ignatius J. Reilly."	High: PRI-CFD-001–005, PRI-CFD-008–012; low: PRI-CFD-006–007

Aesthetic differentiation and reader preference

Aesthetic differentiation was examined across the same twelve synthetic narratives evaluated for coherence, preserving each document code and treating every story as one independent textual unit. Six dimensions captured distinct aspects of literary evaluation: originality, stylistic control, imagery, emotional resonance, interpretive depth, and defamiliarization. Scores ranged from one to five and required textual evidence rather than general impressions of readability. The overall mean reached 2.46, equivalent to 49.2% of the maximum score. Three narratives occupied the high aesthetic range, two were moderate, and seven remained below 2.50. This distribution differs from the predominantly strong coherence profile, indicating that orderly event progression did not automatically produce stylistic distinction, emotional force, or interpretive complexity within this corpus.

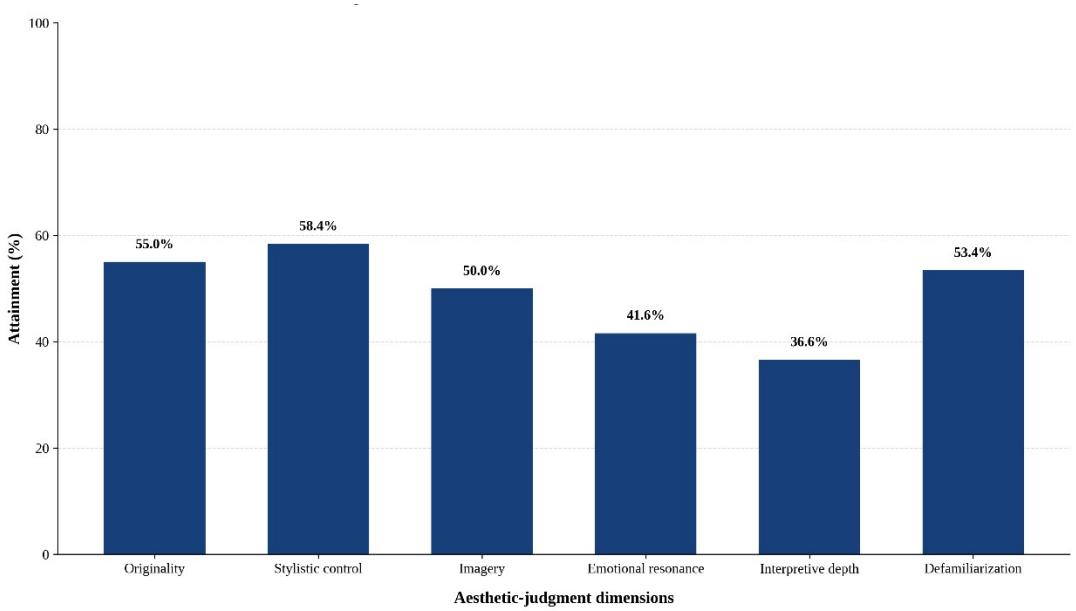


Figure 4. Aesthetic attainment across six dimensions

Stylistic control produced the strongest dimension at 58.3%, followed by originality at 55.0% and defamiliarization at 53.3%. Claude's imperative, "Onward, my scaly steed!" sustained a comic-epic register, while GPT-4's claim that "the very fabric of time and space seemed to rend asunder" combined elevated diction with narrative absurdity. Imagery reached 50.0%, but emotional resonance fell to 41.7%, and interpretive depth recorded only 36.7%. Bing Creative provided a notable exception through "Fortuna has smiled upon me today," connecting action with Ignatius's philosophical preoccupations. Several lower-scoring stories instead ended with generalized lessons about courage, patriotism, or perseverance. Such conclusions completed narrative action but reduced ambiguity, characterization, and opportunities for sustained literary interpretation.

This separation between formal coherence and aesthetic value supports multidimensional approaches to synthetic literature. Manzanarez et al. (2025) argue that AI fiction requires evaluation across stylistic, narrative, and creative criteria rather than through fluency alone, while Marco, Gonzalo, and Fresno-Fernández (2025) demonstrate that textual features and reader profiles generate divergent preferences. Doshi and Hauser (2024) similarly show that generative assistance can improve individual evaluations while narrowing collective diversity. Within this corpus, the comparatively high reader rank assigned to Vicuna despite its low aesthetic profile further indicates that preference cannot be inferred from rubric scores alone. Aesthetic judgment therefore emerges as a relational assessment: textual qualities constrain evaluation, but reader expectations determine which qualities become salient, pleasurable, or legitimate.

Table 3. Distribution of aesthetic-judgment indicators across twelve synthetic stories

Main Category	Aesthetic Indicator	High, n (%)	Moderate, n (%)	Low, n (%)	Mean Score	Attainment	Dominant Variation	Representative Textual Evidence	Data Codes
Formal innovation	Originality of plot, voice, metaphor, or narrative device	3 (25.0%)	3 (25.0%)	6 (50.0%)	2.75/5	55.0%	Innovation was concentrated in stories combining unexpected objects, philosophical allusion, or comic incongruity; half relied on conventional combat structures.	"Let us see if this creature has a sweet tooth."	Strong: PRI-CFD-001, PRI-CFD-004, PRI-CFD-005; low: PRI-CFD-002, PRI-CFD-006, PRI-CFD-008, PRI-CFD-009, PRI-CFD-011, PRI-CFD-012
Linguistic execution	Stylistic control across diction, syntax, tone, and register	3 (25.0%)	6 (50.0%)	3 (25.0%)	2.92/5	58.3%	Most outputs maintained readable prose but varied in tonal consistency; stronger texts sustained comic-epic registers, whereas weaker outputs shifted abruptly into generic moral language.	"Onward, my scaly steed!"	Strong: PRI-CFD-001, PRI-CFD-004, PRI-CFD-005; low: PRI-CFD-006, PRI-CFD-007, PRI-CFD-012
Sensory construction	Imagery and descriptive specificity	2 (16.7%)	2 (16.7%)	8 (66.7%)	2.50/5	50.0%	Vivid imagery appeared selectively, while most narratives depended on generic movement, heroic description, and repetitive combat vocabulary.	"The primordial beast descended from the sky with a reptilian screech."	Strong: PRI-CFD-001, PRI-CFD-004; moderate: PRI-CFD-003, PRI-CFD-005
Affective engagement	Emotional resonance and affective development	0 (0.0%)	4 (33.3%)	8 (66.7%)	2.08/5	41.7%	Emotional force remained limited because conflict was resolved rapidly and characters rarely developed beyond fear, triumph, or heroic self-affirmation.	"Ignatius returned home convinced that he had changed from a slothful person into a warrior."	Moderate: PRI-CFD-001, PRI-CFD-004, PRI-CFD-005, PRI-CFD-010
Semantic	Interpretive depth and	1 (8.3%)	2 (16.7%)	9 (75.0%)	1.83/5	36.7%	Most stories closed with	"Fortuna has	Strong: PRI-

complexity	thematic layering						explicit morals rather than sustained ambiguity; depth increased when philosophical allusion or ironic characterization complicated the battle narrative.	smiled upon me today!"	CFD-005; moderate: PRI-CFD-001, PRI-CFD-004
Perceptual estrangement	Defamiliarization through unusual language, form, or juxtaposition	3 (25.0%)	2 (16.7%)	7 (58.3%)	2.67/5	53.3%	Productive estrangement emerged through anachronism, comic objects, and improbable combinations, although several anomalies functioned as incoherence rather than meaningful innovation.	"Chapter 1—The Immortal Ignatius J. Reilly."	Strong: PRI-CFD-001, PRI-CFD-004, PRI-CFD-005; moderate: PRI-CFD-003, PRI-CFD-007
Composite aesthetic profile	Combined six-dimensional evaluation	3 high (25.0%)	2 moderate (16.7%)	7 low (58.3%)	2.46/5	49.2%	GPT-4, Claude, and Bing formed the high group; ChatGPT Free and Bard occupied the middle range; seven outputs remained below 2.50.	"The very fabric of time and space seemed to rend asunder."	High: PRI-CFD-001, PRI-CFD-004, PRI-CFD-005; moderate: PRI-CFD-003, PRI-CFD-010; low: remaining seven texts

Reader trust, literary legitimacy, and continuation potential

Reader trust was evaluated across the same twelve synthetic stories and retained document codes used in preceding analyses. Because available public data contain story rankings rather than experimental trust responses, coding focused on textual affordances that could support or weaken trust: narrative reliability, perceived authenticity, attributed intentionality, literary legitimacy, continuation potential, and recommendation potential. Each dimension was scored from one to five using linked quotations and analytic memos. Authorship-disclosure sensitivity was excluded from scoring because no blinded–disclosed comparison was available. Composite trust averaged 2.81 out of 5, or 56.1% attainment. Three stories occupied the high range, five were moderate, and four remained low, indicating that coherent narration did not consistently translate into broader literary confidence.

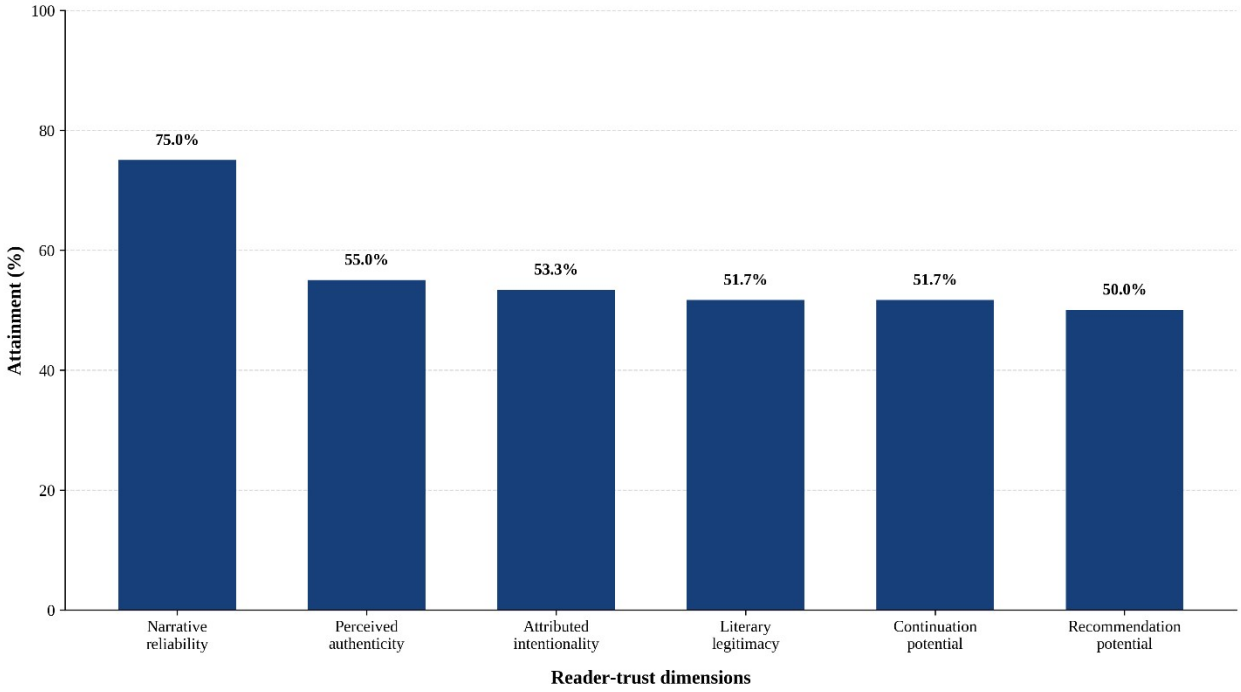


Figure 5. Textually afforded reader trust across six dimensions

Narrative reliability was dominant, reaching 75.0%, with ten stories receiving high scores and only StableLM and Dolly falling into the low range. Bard's sequence Ignatius "repeatedly dodged the creature… until it became tired and flew away" made event consequences easy to follow. Perceived authenticity reached 55.0%, while attributed intentionality achieved 53.3%. Bing's "Fortuna has smiled upon me today!" connected the encounter with Ignatius's philosophical characterization, whereas ChatGPT Free's "Let us see if this creature has a sweet tooth" framed an improbable action as purposeful comedy. Literary legitimacy and continuation potential each reached 51.7%, while recommendation potential was lowest at 50.0%. Generic heroic endings and abrupt identity shifts weakened confidence despite locally intelligible plots.

Trust therefore emerged as a layered judgment rather than a direct consequence of coherence. Ali et al. (2026) associate engagement with coherence, emotional resonance, and authorship perception, while Bajohr (2024) shows that expectations surrounding machine writing shape recognition and value. This corpus supports that distinction: reliable sequencing could invite comprehension, yet limited depth, formulaic closure, or unstable identity reduced authenticity and literary legitimacy. Rodrigues (2026) similarly treats creative authenticity as an epistemological attribution rather than a property that can be verified from stylistic fluency alone. Marco, Gonzalo, and Fresno-Fernández (2025) further demonstrate that reader profiles mediate evaluations of AI writing. Accordingly, textual organization enables trust, but aesthetic purpose and interpretive richness determine whether that trust extends toward literary acceptance.

Table 4. Distribution of textually afforded reader-trust indicators across twelve synthetic stories

Main Category	Trust Indicator	High, n (%)	Moderate, n (%)	Low, n (%)	Mean Score	Attainment	Dominant Variation	Representative Textual Evidence	Data Codes
Interpretive dependability	Narrative reliability: confidence that events, characters, and resolutions can be followed consistently	10 (83.3%)	0 (0.0%)	2 (16.7%)	3.75/5	75.0%	Most stories offered sufficiently stable narrative information, although two texts disrupted reliability through protagonist substitution, temporal rupture, or unresolved restructuring.	"He repeatedly dodged the creature… until it became tired and flew away."	High: PRI-CFD-001–005, PRI-CFD-008–012; low: PRI-CFD-006–007
Expressive credibility	Perceived authenticity: extent to which narrative voice appears distinctive rather than mechanically generic	3 (25.0%)	4 (33.3%)	5 (41.7%)	2.75/5	55.0%	Authenticity increased when diction reflected recognizable characterization or tonal control; standardized heroic language reduced perceived expressive integrity.	"Fortuna has smiled upon me today!"	High: PRI-CFD-001, PRI-CFD-004, PRI-CFD-005; moderate: PRI-CFD-002, PRI-CFD-003, PRI-CFD-008, PRI-CFD-010
Purposeful construction	Attributed intentionality: degree to which narrative choices appear deliberate rather than accidental	3 (25.0%)	4 (33.3%)	5 (41.7%)	2.67/5	53.3%	Comic objects and philosophical references appeared purposeful in stronger texts, whereas abrupt setting, identity, or thematic changes appeared insufficiently motivated.	"Let us see if this creature has a sweet tooth."	High: PRI-CFD-001, PRI-CFD-004, PRI-CFD-005; low: PRI-CFD-006–009, PRI-CFD-012
Cultural-literary acceptance	Literary legitimacy: willingness to treat narrative as an object of sustained literary interpretation	3 (25.0%)	3 (25.0%)	6 (50.0%)	2.58/5	51.7%	Legitimacy was strongest where coherence intersected with stylistic distinction, irony, or thematic allusion; merely complete plots did not secure literary status.	"The very fabric of time and space seemed to rend asunder."	High: PRI-CFD-001, PRI-CFD-004, PRI-CFD-005; moderate: PRI-CFD-002, PRI-CFD-003, PRI-CFD-010

Sustained engagement	Continuation potential: textual capacity to encourage further reading	3 (25.0%)	3 (25.0%)	6 (50.0%)	2.58/5	51.7%	Continuation potential depended on controlled voice, comic momentum, and interpretive promise; repetitive combat and moralizing closure reduced prospective engagement.	"Onward, my scaly steed!"	High: PRI-CFD-001, PRI-CFD-004, PRI-CFD-005; moderate: PRI-CFD-002, PRI-CFD-003, PRI-CFD-010
Public endorsement	Recommendation potential: textual capacity to justify recommending a story as a literary experience	3 (25.0%)	2 (16.7%)	7 (58.3%)	2.50/5	50.0%	Recommendation was the weakest dimension because readable narration often lacked sufficient originality, affective depth, or stylistic memorability.	"The very fabric of time and space seemed to rend asunder."	High: PRI-CFD-001, PRI-CFD-004, PRI-CFD-005; moderate: PRI-CFD-003, PRI-CFD-010
Composite trust profile	Combined six-dimensional textual trust affordance	3 high (25.0%)	5 moderate (41.7%)	4 low (33.3%)	2.81/5	56.1%	GPT-4, Claude, and Bing occupied the high range; Vicuna, ChatGPT Free, GPT4All-J, Bard, and OpenAssistant were moderate; four stories remained low.	"It was the great British patriot, Sir Arthur Seaton."	High: PRI-CFD-001, PRI-CFD-004, PRI-CFD-005; low: PRI-CFD-006, PRI-CFD-007, PRI-CFD-009, PRI-CFD-012

DISCUSSION

The coherence profile matters because synthetic fiction appears capable of sustaining narrative intelligibility without consistently producing robust literary organization. Most stories maintained recognizable event sequences, stable entities, temporal order, and closure, indicating that contemporary generators can satisfy narrative expectations. Yet causal linkage remained comparatively fragile, and breakdowns clustered rather than appearing as isolated defects. This distinction is consequential: readers may follow what happens while remaining uncertain why events occur or why particular actions matter. Callan and Foster (2023) similarly separate coherence from interestingness, whereas Goyal, Li, and Durrett (2022) show that global fluency can conceal local narrative failures. Findings therefore challenge evaluations that treat grammatical continuity as sufficient evidence of story quality. Functionally, coherence provides an entry condition for reading; dysfunctionally, weak causality reduces interpretive confidence and makes closure feel procedural. Synthetic fiction can thus be readable without becoming narratively persuasive, a difference essential for literary evaluation and reader trust.

This pattern is best explained by an asymmetry between sequential generation and long-range narrative planning. Language models can extend locally probable actions, preserve salient entities, and reproduce familiar endings, but causal architecture requires control over motivations, consequences, and thematic relevance. Alabdulkarim, Li, and Peng (2021) identify long-range coherence as a difficulty in automatic story generation, while Papalampidi, Cao, and Kociský (2022) emphasize entity consistency as only one component of narrative control. Li et al. (2025) respond by introducing plot-planning and retrieval mechanisms, suggesting that coherence problems arise from architectural limits rather than lexical incompetence. However, this corpus complicates deterministic accounts: several systems produced strong causal sequences under the same prompt, so breakdown cannot be attributed simply to machine authorship. A more plausible interpretation is that model design, prompting, and learned genre conventions jointly shape narrative stability. Coherence therefore emerges as a structured affordance, not a uniform property of synthetic text.

The aesthetic findings are significant because they reveal that formal competence does not automatically generate literary distinction. Stories could be coherent yet remain weak in emotional resonance, interpretive depth, imagery, or stylistic memorability. This gap matters for digital literary studies because it separates successful narrative assembly from aesthetic achievement in practice. Manzanarez et al. (2025) advocate multidimensional evaluation for AI microfiction, while Marco, Gonzalo, and Fresno-Fernández (2025) demonstrate that reader judgments depend on textual properties and reader profiles. Findings support these approaches by showing that originality, style, and defamiliarization can rise above other dimensions without producing uniformly strong literary value. They also qualify Porter and Machery's (2024) evidence that AI-generated poetry may receive favorable ratings, because positive evaluation in one genre does not erase limitations in another. Synthetic fiction functions aesthetically when formal surprise becomes meaningful; it dysfunctions when novelty remains decorative, formulaic, or disconnected from emotional and thematic development.

The uneven aesthetic profile reflects how generative systems optimize plausibility more readily than significance. Familiar heroic plots, rapid conflict resolution, and explicit moral endings are safe because they resemble familiar narrative patterns, but they compress ambiguity and reduce opportunities for interpretation. Doshi and Hauser (2024) show that generative assistance can improve individual evaluations while narrowing collective diversity, a tension visible in conventional outputs. Rettberg and Wigers (2025) likewise identify stability, homogeneity, and stereotyping in generated stories, whereas Beguš (2024) reports differences in imaginative range between human and machine storytelling. Still, texts complicated this tendency through comic objects, philosophical allusion, and tonal incongruity. Those exceptions indicate that aesthetic weakness is not inevitable but depends on whether unusual choices cohere into recognizable artistic logic. Defamiliarization becomes valuable only when readers can infer purpose. Random anomaly appears accidental; patterned strangeness appears authored. This distinction explains why originality and legitimacy did not move together.

The trust profile carries an implication: readers can rely on synthetic fiction as intelligible discourse without accepting it as literature. Narrative reliability was stronger than authenticity, legitimacy, continuation,

or recommendation potential, showing that trust operates in layers. Comprehension-based trust concerns whether events can be followed; literary trust concerns whether choices deserve interpretation, emotional investment, and endorsement. Bajohr (2024) argues that machine writing transforms readers' expectations of literary and non-literary texts, while Ali et al. (2026) connect engagement with coherence, affect, and authorship perception. Findings extend those arguments by locating distrust not primarily in obvious error but in limited intentionality and depth. This distinction tempers claims that authorship detection determines reception. Readers may accept a story's internal logic while withholding cultural legitimacy. Synthetic fiction therefore enters literary space conditionally: reliability opens access, but authenticity and aesthetic purpose determine whether readers remain, recommend, and treat a text as worthy of sustained attention.

The structure underlying this conditional trust is relational rather than textual. Trust develops when coherent organization, aesthetic differentiation, and attributed purpose reinforce one another; it weakens when one dimension cannot support another. Rodrigues (2026) treats creative authenticity as an epistemological attribution rather than an observable property, and Stańko-Kaczmarek, Dera, and Kościelska (2024) show that authorship labels can influence responses to poetry. Yet this corpus lacks a controlled disclosure comparison, so machine attribution cannot be treated as the cause of trust differences. Instead, observed patterns suggest an association between textual signals and readers' willingness to infer meaningful agency. Marco, Gonzalo, and Fresno-Fernández (2025) similarly emphasize conflicting evaluations across reader profiles, indicating that no feature guarantees acceptance. This study positions reader trust as negotiated judgment: narrative coherence supplies reliability, aesthetic form supplies value, and cultural expectations supply legitimacy. Their convergence makes synthetic literature persuasive; their divergence keeps acceptance provisional, contested, and reader-dependent.

CONCLUSION

This study shows that synthetic fiction can achieve strong narrative coherence without securing equivalent aesthetic value or reader trust. Its central lesson is that readability, literary distinction, and legitimacy are related but nonidentical dimensions. By integrating a narrative coherence matrix, an aesthetic judgment rubric, and a reader trust reception grid across the same twelve texts, this study offers a methodologically aligned framework for evaluating machine-generated literature. Its contribution lies in shifting attention from authorship detection toward layered reader evaluation, refining research questions about how textual organization, stylistic purpose, and perceived intentionality jointly shape acceptance of AI-generated fiction as contemporary literature.

This study is limited by its small corpus, single prompt, English-language focus, reliance on one preserved reader-ranking frame, and absence of a controlled authorship-disclosure experiment. Textual trust scores therefore indicate interpretive affordances rather than direct population-level responses, while model differences cannot be generalized beyond this dataset. Future research should examine larger multilingual corpora, multiple genres, repeated prompts, and diverse reader groups using blinded and disclosed conditions. Longitudinal designs could test whether familiarity with generative literature alters aesthetic criteria over time. Combining close reading, psychometric validation, and experimentally grounded reception data would strengthen claims about coherence, preference, authenticity, and literary legitimacy.

ACKNOWLEDGMENTS

The author thanks the reviewers and editorial team for their constructive feedback on this manuscript.

FUNDING INFORMATION

The author states no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Devita Septian Dwi Hidayati: conceptualization, methodology, formal analysis, data curation, writing – original draft, writing – review and editing (all lead, sole author).

CONFLICT OF INTEREST STATEMENT

The author states no conflict of interest.

AI USE DECLARATION

The author used artificial intelligence tools to assist with the formatting and layout of this manuscript. AI was not used to generate the substantive content, conceptualization, data, analysis, or conclusions of the research, all of which remain the sole responsibility of the author.

INFORMED CONSENT

Not applicable – this study did not involve direct human participants. Data consisted solely of publicly accessible AI-generated short stories and platform metadata.

ETHICAL APPROVAL

This research did not involve human or animal subjects; it analyzed publicly accessible AI-generated short stories and reader-ranking metadata, and therefore no institutional review board approval was required. Where research does involve human subjects, the author affirms that such research would be conducted in full compliance with all relevant national regulations and institutional policies, in accordance with the tenets of the Declaration of Helsinki, and would be approved by the author's institutional review board or equivalent committee.

DATA AVAILABILITY

The corpus/data analyzed in this study are available from the corresponding author upon reasonable request.

REFERENCES

Abdillah, Y. A. (2026). Poetics of algorithmic excess: Digital aesthetics in Indonesia's Twitter poetry bot. Lingua Technica: Journal of Digital Literary Studies, 2(1), 86–101. https://doi.org/10.64595/lingtech.v2i1.138

Alabdulkarim, A., Li, S., & Peng, X. (2021). Automatic story generation: Challenges and attempts. arXiv. https://doi.org/10.18653/v1/2021.nuse-1.8

Ali, E., Ahmed, F., Usmani, S., & Kottaparamban, M. (2026). Readers' engagement with and perception of AI-generated narratives in the context of digital literature. Journal of Language Teaching and Research. https://doi.org/10.17507/jltr.1701.33

Alasmari, J., Alzyoudi, M., Alshehri, M., Alshammari, R., & Aldakan, R. (2025). An automated predictive model for evaluating narrative cohesion in children's stories: A computational linguistic approach considering Gérard Genette's narrative structure theory. International Journal of Adolescence and Youth, 30. https://doi.org/10.1080/02673843.2025.2500527

Bajohr, H. (2024a). On artificial and post-artificial texts: Machine learning and the reader's expectations of literary and non-literary writing. Poetics Today. https://doi.org/10.1215/03335372-11092990

Bajohr, H. (2024b). The deixis of literature: On the conditions for recognizing computers as authors. Orbis Litterarum. https://doi.org/10.1111/oli.12450

Beguš, N. (2024). Experimental narratives: A comparison of human crowdsourced storytelling and AI storytelling. Humanities and Social Sciences Communications, 11. https://doi.org/10.1057/s41599-024-03868-8

Callan, D., & Foster, J. (2023). How interesting and coherent are the stories generated by a large-scale neural language model? Comparing human and automatic evaluations of machine-generated text. Expert Systems, 40, e13292. https://doi.org/10.1111/exsy.13292

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10, eadn5290. https://doi.org/10.1126/sciadv.adn5290

Fawaid, A. (2025). Mapping the field of digital literary studies: Concepts, methods, and emerging directions. Lingua Technica: Journal of Digital Literary Studies, 1(1), 1–13. https://doi.org/10.64595/99c6t274

Goyal, T., Li, J. J., & Durrett, G. (2022). SNaC: Coherence error detection for narrative summarization. arXiv. https://doi.org/10.48550/arXiv.2205.09641

Li, J., Chen, Y., Liu, Z., Tan, M., Zhang, L., Li, Y., Luo, R., Chen, L., Luo, J., Argha, A., Alinejad-Rokny, H., Zhou, W., & Yang, M. (2025). STORYTELLER: An enhanced plot-planning framework for coherent and cohesive story generation. arXiv. https://doi.org/10.48550/arXiv.2506.02347

Manzanarez, G. A., De La Cruz Arana, N., Flores, J., Medina, Y. G., Monroy, R., & Pernelle, N. (2025). Can artificial intelligence write like Borges? An evaluation protocol for Spanish microfiction. Applied Sciences, 15(12), 6802. https://doi.org/10.3390/app15126802

Marco, G., Gonzalo, J., & Fresno-Fernández, V. (2025). The reader is the metric: How textual features and reader profiles explain conflicting evaluations of AI creative writing. In Findings of the Association for Computational Linguistics: ACL 2025. https://doi.org/10.48550/arXiv.2506.03310

Mikros, G. (2025). Beyond the surface: Stylometric analysis of GPT-4o's capacity for literary style imitation. Digital Scholarship in the Humanities, 40, 587–600. https://doi.org/10.1093/llc/fqaf035

Nor, M. R. M., & Zubaidi, A. (2026). Digital reading platforms in literary learning: Cultural understanding and engagement in Indonesian and Malay texts. Lingua Technica: Journal of Digital Literary Studies, 2(1), 33–50. https://doi.org/10.64595/lingtech.v2i1.131

Papalampidi, P., Cao, K., & Kočiský, T. (2022). Towards coherent and consistent use of entities in narrative generation. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (pp. 17278–17294).

Piper, A., So, R., & Bamman, D. (2021). Narrative theory for computational narrative understanding. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (pp. 298–311). https://doi.org/10.18653/v1/2021.emnlp-main.26

Porter, B., & Machery, E. (2024). AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably. Scientific Reports, 14. https://doi.org/10.1038/s41598-024-76900-1

Rettberg, J. W., & Wigers, H. (2025). AI-generated stories favour stability over change: Homogeneity and cultural stereotyping in narratives generated by GPT-4o-mini. Open Research Europe. https://doi.org/10.12688/openreseurope.20576.1

Rodrigues, T. V. (2026). The epistemology of algorithmic narrative and the problem of creative authenticity. Philosophy & Technology, 39. https://doi.org/10.1007/s13347-025-01011-2

Stańko-Kaczmarek, M., Dera, L., & Kościelska, H. (2024). "Between the lines": Perceptions of poetry with authorship attributed to artificial intelligence or humans—A comparative analysis. The Journal of Creative Behavior. https://doi.org/10.1002/jocb.1513

Suaidi. (2025). Reader participation, platform affordances, and interactive meaning-making in online literary environments. Lingua Technica: Journal of Digital Literary Studies, 1(1), 54–66. https://doi.org/10.64595/9gy6p728

Wang, S., & Huang, G. (2024). The impact of machine authorship on news audience perceptions: A meta-analysis of experimental studies. Communication Research, 51, 815–842. https://doi.org/10.1177/00936502241229794

Yang, D., & Jin, Q. (2024). What makes a good story and how can we measure it? A comprehensive survey of story evaluation. arXiv. https://doi.org/10.48550/arXiv.2408.14622