Humans Introduce, Models Elaborate: Asymmetric Narrative Agency in Human–LLM Co-Writing

Anonymous ACL submission

Abstract

Human–LLM co-writing is increasingly used for open-ended text generation, but much prior work focuses on final outputs rather than the interactional dynamics through which stories are produced. We study turn-based collaborative storytelling across three matched conditions: Human–Human (HH), Human–LLM (HA), and LLM–LLM (AA). Using a shared storytelling paradigm, we measure how agents align, introduce novel material, and influence narrative development through turn-level measures of valence adaptation, semantic novelty, transience, and resonance. Our results show that HA co-writing is not intermediate between HH and AA collaboration. Instead, it displays a distinctive asymmetry where humans tend to introduce more novel and persistent narrative material, while LLMs tend to elaborate and stabilize the existing context. These findings suggest that, in this setting, LLMs function less as human co-authors and more as adaptive narrative amplifiers that reshape how agency is distributed in collaborative writing.

1 Introduction

Large language models (LLMs) are increasingly used as partners in open-ended writing tasks, from ideation and revision to full narrative co-creation (Noy and Zhang, 2023; Clark and Smith, 2021; Cheng and Zhang, 2025). In such settings, the relevant question is how the text came to be. A co-written story unfolds through a sequence of turns: each contribution responds to prior context, shifts the semantic and emotional trajectory of the story, and constrains what can come next. Understanding Human–LLM co-writing therefore requires methods that model collaboration as a dynamic process. Turn-based storytelling provides a controlled setting for studying this process. Existing work on Human–LLM co-creation has focused on output quality, creativity, productivity and user experience. These perspectives are important, but leave open an interactional question: does Human–LLM co-writing resemble Human–Human collaboration, LLM–LLM collaboration, or a distinct regime of narrative coordination? Answering this question requires direct comparison of different dyadic pairings on the same experimental paradigm. Figure 1 summarizes the shared setup and analysis pipeline. We address this gap by comparing three turn-based co-writing conditions: **Human–Human (HH), Human–LLM (HA)**, and **LLM–LLM (AA)**. Building on an existing Human–LLM corpus, we introduce matched Human–Human and LLM–LLM conditions, allowing us to compare collaborative dynamics while holding the task structure constant. We analyze each story at the turn level using measures of affective alignment and semantic influence. This comparative design lets us test whether asymmetry is a general property of dyadic storytelling or a specific feature of HA collaboration. We treat narrative agency as a measurable distribution of local influence, addressing who shifts the story, whose contributions are taken up, and how strongly partners adapt to each other.

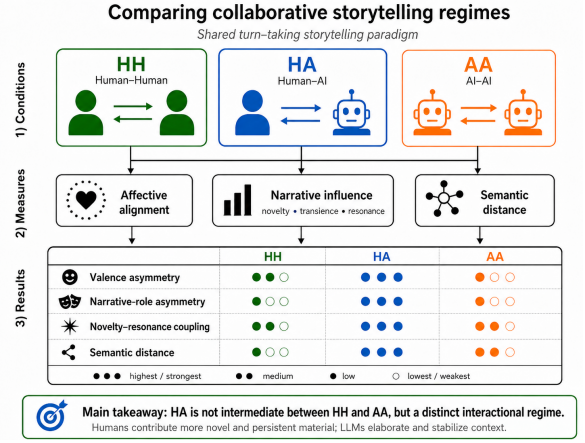

Figure 1: **Overview of setup and pipeline.** Stories are generated under HH, HA, and AA conditions, then analyzed for valence alignment, narrative influence, and semantic distance.

We investigate three research questions:

1. **RQ1:** How do affective alignment and semantic distance differ across HH, HA, and AA co-writing?
2. **RQ2:** How does narrative influence distribute across HH, HA, and AA co-writing?
3. **RQ3:** Does HA constitute a distinct interactional regime or an intermediate condition between HH and AA?

Our contributions are threefold. First, we construct a matched comparative dataset of HH, HA, and AA turn-based stories. Second, we apply a turn-level framework for measuring affective adaptation and directional narrative influence in collaborative writing. Third, we show that Human–LLM co-writing exhibits its clearest asymmetry in semantic influence: humans tend to contribute more novel and persistent material, while LLMs tend to elaborate the evolving narrative context. This suggests that LLMs do not simply replace human authors, but redistribute narrative agency in a unique way.

2 Related Work

Human–LLM co-creation Human–LLM co-writing holds the promise of hybrid intelligence: the possibility that human and artificial writers may collaborate to produce new forms of creativity (Cheng and Zhang, 2025). Much prior work on human–LLM creativity evaluates output and task-level outcomes such as quality, originality, productivity, or ideation, with mixed findings. Domanti et al. (2026) compared human and LLM creativity by relating semantic network diversity to originality, showing that LLM-generated outputs may be more original than outputs of low-creative humans. In specific convergent and divergent thinking tasks, LLMs have been shown to outperform student participants (Arora et al., 2025). Human–LLM co-creation may furthermore increase the absolute number of novel artifacts (Zhou et al., 2025), task enjoyment, and writing quality (Noy and Zhang, 2023). On the other hand, co-creation can narrow content diversity, homogenizing ideas and style (Wenger and Kenett, 2025; Anderson et al., 2024). Recent work suggests that human–LLM joint creativity may reduce participants' self-perceived creativity (Luchini et al., 2025) and, in some settings, may underperform human–human collaboration on creative outcomes (Tang et al., 2025).

Alignment and affective adaptation A second line of work studies alignment and adaptation in dialogue, showing that human and artificial agents exhibit forms of linguistic and emotional coordination. Such coordination is observable on semantic, affective, syntactic, phonetic and lexical levels, and can support effective communication (Pickering and Garrod, 2004; Branigan et al., 2010). Koulouri et al. (2016) showed how lexical alignment occurs in human–computer exchanges, through vocabulary adaptation. Other lexical approaches have revealed differences between human–human and human–agent dialogue (Duplessis et al., 2021; Li et al., 2025), and have shown consistent human alignment to artificial agents (Ostrand et al., 2023). However, lexical matching captures one limited dimension of coordination, and may overlook the underlying emotional contagion that unfolds during dialogue (Hatfield et al., 1993). On this account, Varni et al. (2017) proposed a computational framework for assessing emotional contagion by matching time series polarity states between interlocutors, while other approaches have modeled it by categorizing utterances into predefined emotion labels (Poria et al., 2019).

Sentiment analysis in digital humanities For modeling affective dynamics, computational sentiment analysis has emerged as a valuable methodological foundation (Mohammad, 2016). Prior work has modeled emotional arcs and sentiment progression in literary narratives (Reagan et al., 2016; Hu et al., 2021; Öhman and Rossi, 2022). However, dictionary-based approaches have weak context awareness, motivating the use of transformer-based alternatives (Feldkamp et al., 2024). Both methods remain challenging to apply as general-purpose sentiment models to literary text. Lyngbaek et al. (2026) addressed this by using concept vector projection as a method for deriving continuous sentiment scores across domains and languages. Building on this, Fundal and Bizzoni (2026) adopted concept vector projection to model sentiment alignment between human–LLM interlocutors in co-creative writing.

Information propagation in static discourse and dialogue A central question in discourse analysis concerns how contributions diverge from prior context and whether they shape what follows. Barron et al. (2018) introduced an information-theoretic framework for measuring novelty, transience, and resonance in large-scale corpora of French Revo-

Condition	Stories	Exchanges	Turns	Mean words
HH	36	360	720	26.13
HA	97	900	1800	26.96
AA	80	800	1600	28.24

Table 1: Dataset statistics across co-writing conditions. *Turns* are individual agent contributions; *exchanges* are adjacent pairs we use for alignment and influence analyses.

lution speeches. In their paradigm, novelty captures departure from prior discourse, transience captures limited future uptake, and resonance encompasses contributions that are both novel and persistent. Extending these information-theoretic intuitions to conversational paradigms, Bergey and DeDeo (2024) used LLM raw surprisal scores of utterances to model information flow and to predict backchannels in dialogue, and Maës et al. (2022) used a similar approach to show how the stability of surprisal-based information varies locally within themes. Other work on dialogue has shown how information density converges between interlocutors over time (Xu and Reitter, 2018). However, how these information-propagation measures operate in co-creative human–LLM narratives remains less explored.

3 Datasets

We use a turn-based storytelling paradigm across three dyadic conditions to compare collaborative dynamics while holding a constant task structure.[1]. Participant recruitment, preprocessing, and demographics are in Appendix A.

3.1 Shared task design

Across conditions, agents received the same collaborative storytelling instruction:

> *"You are an author taking part in a collaborative storytelling activity with another author. Together, you will create a story by taking turns adding to it. Your goal is to continue from where your partner has left off. If there is no story, please begin the story. You have 10 exchanges to write the story."*

We designed the prompt to elicit open-ended narrative continuation. The fixed turn-taking format makes each story analyzable as a sequence of local responses and allows us to measure how each agent adapts to the preceding contributions.

Our main analyses distinguish between individual turns and directed exchanges. **Turns** are individual agent contributions, and **directed exchanges** are adjacent turn pairs we use to measure response, alignment, and influence. We aggregate turn-level metrics by story for cross-condition comparison.

3.2 Human–LLM condition

The Human–LLM data comes from Fundal and Bizzoni (2026). In this condition, human participants alternated turns with an LLM in the collaborative storytelling interface. There were 97 participants (male = 34, female = 40, other = 4) with an age distribution of M = 27.34 years (SD = 14.66). [2]

3.3 Human–Human

The HH condition used the same turn-taking structure as the HA condition and included 72 participants (male = 27, female = 32, other = 2) with an age distribution of M = 26.98 years (SD = 7.16). Participants were university or graduate students at Aarhus University from English-taught programs such as Linguistics, English, Cultural Data Science and Cognitive Science. Participation was voluntary and uncompensated. The platform paired participants anonymously. This dynamic produced a comparable human baseline for turn-based narrative collaboration.

3.4 AA condition

The AA condition replicates the same turn-taking constraints in a simulation pipeline. For each LLM used in the HA condition, two agents instantiated as the same model generated new stories, allowing us to compare HA and HH collaboration with a model-model (AA) interaction under matched structural constraints.

3.5 Preprocessing and filtering

We removed incomplete sessions, empty turns, and malformed outputs before analysis. Because our measures depend on adjacent turns, any missing contribution affects the number of valid exchange-level observations. Some HA stories only had 9 full exchanges, resulting in slightly fewer interactions relative to the number of stories. All reported

[1] We will release our dataset and analysis framework upon acceptance

[2] The LLMs used in this condition and in our AA condition were `gpt-4.1-2025-04-14`, `claude-sonnet-4-5-20250929`, `Llama-3.3-70B-Instruct`, and `Qwen-2.5-72B-Instruct`.

statistics in Table 1 refer to the final analytic dataset used in the experiments.

4 Methods

Our goal is to quantify how collaborative storytelling distributes labor across turns. We compute turn- and exchange-level measures capturing three complementary aspects of co-writing: (i) affective alignment, revealing how agents adapt to the previous turns' emotional tone; (ii) semantic distance, revealing how similar adjacent contributions are; (iii) narrative influence, revealing whether a contribution introduces material that shapes the following turn. We treat narrative agency as the degree to which each agent introduces material that diverges from context (novelty), whether that material is taken up by what follows (resonance), and how strongly partners track each other's valence and semantic trajectories (alignment).

4.1 Baseline valence difference

For sentiment computation, we use concept vector projection, following the approach introduced by (Lyngbaek et al., 2026). For each turn T_t we compute an embedding e_t and project it onto a pretrained sentiment concept vector $\hat{c}$. Turn embeddings were computed using `paraphrase-multilingual-mpnet-base-v2`, and all projections were computed on normalized embeddings. The resulting scalar score,

$$v_t = e_t \cdot \hat{c}$$

represents the position of the turn on a negative–positive valence axis. Lower values indicate negative affective tones, higher values indicate positive tones.

Because the HH and AA conditions do not contain a privileged agent role, our main cross-condition analyses use unsigned asymmetry magnitudes. This allows us to ask whether a condition produces more imbalance between partners, independently of which side is larger.

For baseline differences, we assess the absolute difference of mean valence scores between agents per story.

$$A_{\text{baseline}}(S_i) = \left|\bar{v}_i^{(1)} - \bar{v}_i^{(2)}\right| \tag{1}$$

where $\bar{v}_i^{(a)}$ is the mean valence of agent a in story S_i. This represents a measure of how far apart the two agents' average valence baselines are within a story.

4.2 Directional Alignment

For asymmetries in directional alignment, we build on the framework from (Fundal and Bizzoni, 2026) by regarding emotional adaptation as a relationship between predictor-turn valence and response-turn valence.

Let $A_{i,t}^{(a)}$ denote the turn produced by agent $a \in \{1, 2\}$ at exchange t in story S_i, and let $v(A_{i,t}^{(a)})$ be its valence score. We estimate directional alignment from agent 1 to agent 2 as the within-story correlation between agent 1's turns and the following responses by agent 2:

$$r_i^{(1\to 2)} = \text{corr}\left(v_{i,t}^{(1)}, v_{i,t}^{(2)}\right)$$

We estimate the reverse direction from agent 2 to agent 1 as the correlation between agent 2's turns and agent 1's following responses. Because turns strictly alternate, each predictor turn is immediately followed by its partner's response. Agent 2's turn at exchange t follows agent 1's turn at t, while agent 1's turn at $t+1$ follows agent 2's turn at t:

$$r_i^{(2\to 1)} = \text{corr}\left(v_{i,t}^{(2)}, v_{i,t+1}^{(1)}\right).$$

For each story, we compute one alignment asymmetry magnitude as the difference in correlation of directional alignment between agents. Since HH and AA conditions have arbitrary directions, we use absolute values.

$$A_{\text{align}}(S_i) = \left|\text{atanh}\left(r_i^{(1\to 2)}\right) - \text{atanh}\left(r_i^{(2\to 1)}\right)\right| \tag{2}$$

where $r_i^{(1\to 2)}$ is the within-story correlation of the directional pairs $(A_{i,t}^{(1)}, A_{i,t}^{(2)})$ and $r_i^{(2\to 1)}$ is the corresponding correlation of the directional pairs $(A_{i,t}^{(2)}, A_{i,t+1}^{(1)})$.

Affective asymmetry is then a measure of the difference between how much each agent immediately adapts to its counterpart.

As an additional measure, we conduct a similar signed within-condition analysis for the HA condition as a mixed-model with an interaction effect for agent difference. Positive interaction values indicate stronger LLM alignment to previous human valence than human alignment to previous LLM valence.

4.3 Semantic distance

We evaluate asymmetry on a semantic level as the cosine distance between each adjacent turn between agents. Since turns are alternating, each

distance value D_t represents local semantic dissimilarity between the two agents in an interaction.

$$D_t = 1 - \frac{e_t^{(1)} \cdot e_t^{(2)}}{\|e_t^{(1)}\| \, \|e_t^{(2)}\|} \tag{3}$$

We model semantic distance as a mixed-effects model with adjacent semantic distance as outcome, condition as fixed effect, and a story-level random intercept.

$$D_{ij} \sim \texttt{condition} + (1 \mid \texttt{conversation_id}) \tag{4}$$

4.4 Novelty, transience and resonance

For narrative influence, we use information-theoretic measures of discourse influence, namely Novelty, Transience and Resonance, adapted from Barron et al. (2018) to a surprisal-based metric. We define each metric as excess surprisal, where more negative values indicate that the conditioning information reduced surprise further. Values closer to zero thus correspond to the high end of the Barron et al. intuition. Novelty measures the divergence of a turn from prior context, Transience measures the surprise effect of a turn on the immediately following story contribution, and Resonance measures the relationship between the two, expressing whether that departure remains predictive of what follows.

Surprisal We build on the theories of surprisal-based novelty, transience, and resonance from Fundal and Bizzoni (2026) and expand the methodological paradigm. The single surprisal score of a token w_j is computed as the negative log-probability of w_j given the preceding tokens in a defined context window $w_{<j}$. The scores were computed using the open-source base model `google/gemma-4-31b`.

$$s(w_j) = -\log_2 p(w_j \mid w_{<j}) \tag{5}$$

The surprisal score of turn-level inputs $\bar{s}(T_t \mid C_t)$ is then computed as the mean surprisal score of the tokens in a turn T_t. Each token is conditioned on the preceding context C_t and the preceding tokens within the turn $w_{t,<j}$.

$$\bar{s}(T_t \mid C_t) = \frac{1}{n_t} \sum_{j=1}^{n_t} -\log_2 p(w_{t,j} \mid C_t, w_{t,<j}) \tag{6}$$

Novelty We then define Novelty and Transience by calculating excess surprisal of the current turn given prior context, relative to a context-free baseline. This can be interpreted as the negative of a model-based PMI estimate between the preceding context and the current turn. The contextual score is computed using all preceding tokens in the story C_t. The baseline score is calculated as the surprisal score of the turn from the model's beginning-of-sequence token (BOS). Formally:

$$\text{Novelty}_t = \bar{s}(T_t \mid C_t) - \bar{s}(T_t \mid \text{BOS}) \tag{7}$$

Less negative novelty indicates relatively greater contextual surprise.

Transience Instead of regarding transience only as the isolated total predictive utility of a turn T_t, regardless of the prior context, we instead measure the marginal contribution of T_t given the previous context for predicting subsequent turns. This is obtained by conditioning on the past and asking how much the current turn increases surprise of the future turn beyond what has already been established. Formally, we transition to

$$\text{Transience}_t = \bar{s}(F_{t+1} \mid C_t, T_t) - \bar{s}(F_{t+1} \mid C_t) \tag{8}$$

where F_{t+1} denotes the immediately following counterpart turn. This facilitates a measure of how the turn carries into the future narrative, in addition to what has already been established. Finally, we define resonance as the net gain of novelty on future discourse.

$$\text{Resonance}_t = \text{Novelty}_t - \text{Transience}_t \tag{9}$$

Resonance is thus an expression of how much a contribution sticks in the future relative to how surprising it is. Stronger future uptake increases resonance, and weaker uptake dampens it. For example, if a turn unexpectedly introduces a dragon and the following turn integrates that dragon, the first turn will be relatively novel while also reducing transience, due to high future uptake. Under our sign convention, this yields stronger resonance.

Modeling We model differences in agent influence asymmetry between conditions on two levels. The first level is the absolute magnitude of asymmetry for each metric $\overline{m}$ (novelty, transience, and resonance).

$$A_m(S_i) = \left| \bar{m}_i^{(1)} - \bar{m}_i^{(2)} \right| \tag{10}$$

where $m \in \{\text{Novelty}, \text{Transience}, \text{Resonance}\}$ and $\bar{m}_i^{(a)}$ is the mean value of metric m for agent a in story S_i.

The second level models novel narrative influence as resonance predicted by novelty, with condition as an interaction term. A novelty-to-resonance slope higher than 1 indicates an innovation bias, where more surprising turns have a relatively stronger influence on the subsequent narrative.

$$\texttt{resonance} \sim \texttt{novelty} * \texttt{condition} + (1 \mid \texttt{conversation_id}) \quad (11)$$

Within HA analysis Since agent identity is meaningful in the HA condition, we additionally report the within-condition agent contrasts for novelty, transience, and resonance, in order to interpret the direction of the unsigned asymmetry measures.

Finally, within HA, we fit an agent-specific novelty-resonance mixed model to test agent differences in innovation-bias. The agent×novelty interaction estimates whether one agent shows a stronger innovation bias than the other. This complements the story-level slope measure by identifying the direction of the agent difference within the HA condition.

$$\texttt{resonance} \sim \texttt{novelty} * \texttt{agent} + (1 \mid \texttt{conversation_id}) \quad (12)$$

5 Results

5.1 Affective alignment

HA stories show larger baseline affective asymmetry HA exhibited the largest difference in average valence. Baseline valence asymmetry was higher in HA stories (mean $A_{\text{baseline}} = 0.060$, 95% CI $[0.051, 0.069]$) than HH (mean $= 0.041$, 95% CI $[0.032, 0.052]$) and AA (mean $= 0.028$, 95% CI $[0.024, 0.033]$). The planned contrast comparing HA against the two same-type baselines was significant ($\Delta = 0.025$, 95% CI $[0.014, 0.036]$, $p < .001$). This suggests that HA collaboration produces a larger affective gap between partners (Figure 2); we test directional adaptation below.

Directional valence alignment shows weak evidence of asymmetry Directional valence alignment was positive under all conditions, indicating agents responded to the affective tone of their partners. However, evidence for cross-condition differences in unsigned alignment asymmetry was weak.

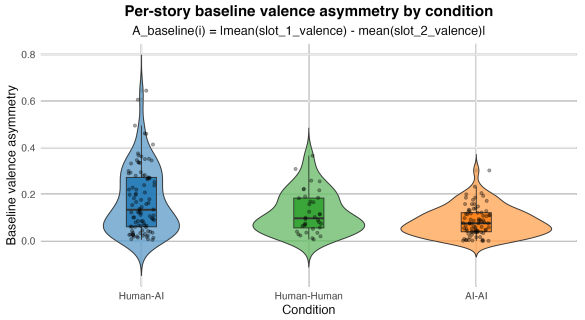

Figure 2: Per-story baseline valence asymmetry by condition.

Condition	n	Mean A_{val}	95% CI	Median
HA	97	0.608	[0.518, 0.705]	0.530
HH	36	0.521	[0.399, 0.646]	0.426
AA	80	0.496	[0.418, 0.578]	0.429

Table 2: Per-story valence alignment asymmetry across conditions. A_{val} is the absolute difference between Fisher-transformed directional valence correlations.

The unsigned per-story alignment-asymmetry measure was highest in HA (mean $A_{\text{valence}} = 0.608$, 95% CI $[0.518, 0.705]$), compared with HH (mean $= 0.521$, 95% CI $[0.399, 0.646]$) and AA (mean $= 0.496$, 95% CI $[0.418, 0.578]$), but the cross-condition tests were not reliable (ANOVA $p = .190$; Kruskal–Wallis $p = .365$; planned HA versus same-type contrast $p = .106$). Table 2 summarizes the unsigned alignment-asymmetry estimates.

The HA directional alignment analysis showed a significant agent-level difference. The mixed-model directional slope gap showed stronger LLM alignment than human alignment ($\Delta = 0.143$, $p = .008$). A signed per-story diagnostic showed the same direction (mean signed $\Delta z = 0.194$, $p = .012$), while corresponding signed slot tests were not reliable for HH or AA.

5.2 Semantic distance

HA exchanges show high semantic distance Although cross-condition differences in affective alignment were weak, semantic distance showed a clearer difference. Adjacent turn-to-turn semantic distance differed by condition. HA stories showed the highest local semantic distance (mean $= 0.316$), followed by AA (mean $= 0.296$) and HH (mean $= 0.287$). In the mixed model, HA distances were larger than HH ($\Delta = 0.0299$, $p < .0001$) and AA ($\Delta = 0.0198$, $p < .0001$). The planned contrast comparing HA with the average of the two same-type conditions was significant ($\Delta = 0.0248$,

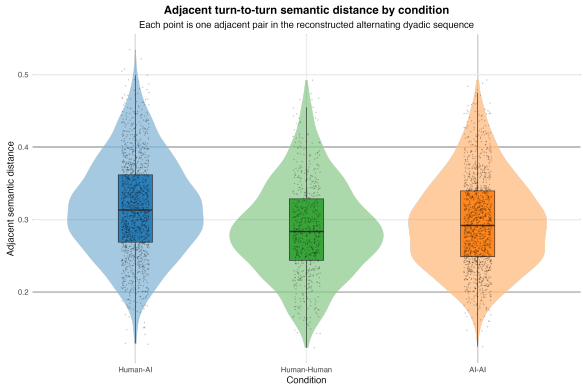

Figure 3: Adjacent semantic distance by condition.

$p < .0001$). This supports the notion that HA co-writing involves greater local semantic asymmetry between adjacent turns (Figure 3).

5.3 Narrative influence

HA co-writing produces the largest narrative-role asymmetry The strongest cross-condition effect appears in metric-level slot asymmetry. HA stories show larger differences between the two writer slots than same-type baselines across all three narrative-influence metrics: novelty (mean $A_m = 0.828$, 95% CI $[0.732, 0.926]$), transience (mean $A_m = 0.484$, 95% CI $[0.432, 0.537]$), and resonance (mean $A_m = 1.290$, 95% CI $[1.149, 1.432]$). Planned contrasts comparing HA with the average of HH and AA were significant for all three metrics ($p_{\mathrm{BH}} < .001$). This result indicates that HA co-writing is different from an intermediate level between HH and AA collaboration: it produces a unique asymmetry in the allocation of narrative functions (Table 3).

Novelty couples with resonance in HA stories Across all conditions, more novel turns tend to be more resonant, indicating that surprising contributions are more likely to shape the following narrative context. This novelty–resonance coupling is strongest in HA stories (slope $= 1.217$, 95% CI $[1.184, 1.250]$). HH stories (slope $= 1.066$, 95% CI $[1.003, 1.128]$) and AA stories (slope $= 0.956$, 95% CI $[0.913, 0.999]$) follow. Both HH and AA slopes were flatter than the HA slope (Figure 4).

Humans contribute higher novelty and resonance in HA stories. Within HA, the agent metric distributions clarify the direction of the asymmetry results (Figure 5). Human turns had higher mean novelty (-1.818 vs. -2.640) and higher mean resonance (-0.796 vs. -2.07) than LLM

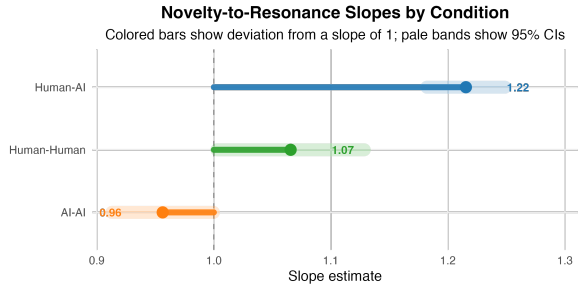

Figure 4: Novelty–resonance relationship by condition.

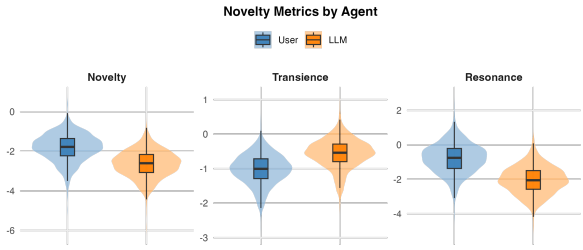

Figure 5: Novelty, transience, and resonance distributions by agent in the HA condition.

turns, while LLM transience was higher (closer to zero: -0.566 vs. -1.022). Under the sign convention introduced in Methods, this means humans introduce more contextually surprising material that is more strongly taken up by the LLM than the other way around. Novelty predicted resonance for both agents, descriptively more strongly for humans, but the agent$\times$novelty interaction was not significant ($\beta_3 = -0.058$, $p = .083$, LLM relative to human reference level). Table 4 summarizes the results by research question.

6 Discussion

Treating co-writing as an unfolding interaction, our work has provided a finer-grained account of creativity and coordination in open-ended text generation. By investigating systematic patterns in role differentiation across all three conditions, we identify differentiation as most pronounced in human–LLM collaboration.

Adaptation of emotional tone appears to be a general dynamic of co-creative storytelling. This is consistent with the idea that accommodation is necessary for producing a coherent narrative. The cross-condition differences do not lie in whether partners adapt, but in the direction and unevenness of that adaptation. Human and LLM contributions maintain distinct affective baselines, and the within-condition HA analysis suggests that the model does most of the accommodation while humans retain

Metric	HA	HH	AA
Novelty	0.828	0.355	0.209
Transience	0.484	0.300	0.197
Resonance	1.290	0.605	0.313

Table 3: Mean metric-level slot asymmetry A_m across conditions. HA exceeds both baselines for all metrics; all planned HA-vs.-baseline contrasts are significant ($p_{\mathrm{BH}} < .001$). Full CIs and Cliff's δ are reported in Table 5.

RQ	Finding	Evidence
RQ2′	Larger HA slot asymmetry in novelty, transience, resonance	Strong
RQ1	Largest adjacent semantic distance in HA	Strong
RQ2″	Novelty–resonance slope steepest in HA	Moderate
RQ1	Largest baseline valence asymmetry; LLMs align more within HA	Moderate
RQ1	Cross-condition alignment and agent×novelty interaction	Trend
RQ3	HA is a distinct regime, not an HH–AA midpoint	—

Table 4: Summary of findings, ordered by evidential strength. Significance values are reported in-text.

more emotional autonomy. Evidence for stronger cross-condition differences was weaker. The directional imbalance between conditions was suggestive but not statistically reliable, so this pattern should be treated as directional evidence rather than an established cross-condition effect. The results on semantic distance and narrative influence are where the HA condition becomes its own interactional regime. The significantly larger adjacent semantic distance in HA highlights how human and LLM contributions occupy more distinct semantic regions of meaning than same-type partners do. The asymmetry results on novelty, transience, and resonance elaborate on the nature of this increased distance. Humans contribute more material that is both surprising and taken up by what follows. In contrast, LLMs seem to consolidate the existing narrative direction, rather than steering it. This constitutes the core asymmetry of role distribution in HA co-writing and is not observed as strongly in the human–human condition and even less so in the LLM–LLM condition. The results confirm that this is a partner-based property, and not a general feature of dyadic storytelling. The novelty-resonance slope-level differences add a second nuanced layer to this pattern. When ignoring partner asymmetry and looking at general novelty-resonance coupling for each condition, HA interactions show a distinctly steeper slope, indicating stronger downstream integration of surprising contributions into the narrative. We interpret this as a stronger innovation bias in the mixed pairing, where more surprising turns are rewarded through future integration, whereas AA interactions seem to stratify surprising content by reduced future integration. Together, these findings support the claim that Human–LLM co-writing is a distinct interactional regime characterized by asymmetric agency and adaptive alignment, with properties that differ from both human–human and LLM–LLM co-writing.

7 Conclusion and Future Work

This work examines turn-based collaborative storytelling as a dynamic interactional process, comparing Human–Human, Human–LLM, and LLM–LLM conditions using a shared experimental paradigm. Our findings suggest that Human–LLM co-writing is not merely an intermediate state, but rather is qualitatively distinct from HH and AA writing, characterized by unique and systematic asymmetries in affective alignment, semantic distance, and narrative influence. The clearest asymmetry lies in narrative agency: humans tend to introduce contextually surprising material that persists into subsequent turns, while LLMs tend to elaborate on and stabilize the emerging narrative context. LLM contributions, however, are not entirely derivative. Rather, they tend to take on an elaborative but active role, exhibiting non-zero levels of novelty, transience, and semantic dissimilarity. Future work should explore whether these observed asymmetric dynamics hold across longer interaction horizons, more diverse participant populations, and different writing genres. Extending the framework to other dyadic text-based settings – such as argument, negotiation, or collaborative explanation – would help establish how general these interactional asymmetries are. In particular, the observed asymmetry in narrative agency leaves us with a set of important open questions: What kind of influence do LLMs exert when they elaborate? Do they steer narratives toward latent structural or archetypal patterns encoded in their representations? Given the critical role that narratives play in shaping our way of thinking, it is important to better understand the subtle but significant influence that LLM co-writers may have.

Limitations

The three conditions differ in sample size, which may affect the reliability of cross-condition comparisons. Generalizability is limited by the fact that participants are predominantly students from a single university. The HA and AA conditions pool stories across multiple LLM families, which increases ecological validity but may obscure model-specific dynamics. The paradigm is also limited to ten exchanges, potentially missing longer-horizon collaborative writing dynamics.

Our measures of valence, novelty, transience, and resonance depend on specific model choices. While absolute values may vary with different tools, we expect relative patterns to be robust. Narrative agency is operationalized through measurable discourse influence rather than direct observation, so our metrics cannot shed light on authorial intention or subjective sense of agency. The experimental setup introduces asymmetries between humans and LLMs that same-pair conditions partially account for by nature. Finally, spell-correction applied to human contributions may have normalized surface text toward LLM-like writing, potentially attenuating some human–LLM lexical differences.

Ethics Statement

This study was conducted according to ethical guidelines. All participants provided consent prior to participation and were informed that their text input would be used for research purposes. Only anonymized demographic data (age, gender) was collected. Participants were free to withdraw at any time. In the human–LLM condition, participants were aware that their co-writer was an LLM. In the human–human condition, participants were paired anonymously with another human participant. The dataset consists of English-language creative fiction produced by adult participants recruited from a university context in Denmark.

References

Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization effects of large language models on human creative ideation. In *Proceedings of the 16th conference on creativity & cognition*, pages 413–425.

Vikram Arora, Alex Thabane, Sameer Parpia, Goran Calic, and Mohit Bhandari. 2025. Generative artificial intelligence models outperform students on divergent and convergent thinking assessments. *Scientific Reports*, 15(1):36987.

Alexander TJ Barron, Jenny Huang, Rebecca L Spang, and Simon DeDeo. 2018. Individuals, institutions, and innovation in the debates of the french revolution. *Proceedings of the National Academy of Sciences*, 115(18):4607–4612.

Claire Augusta Bergey and Simon DeDeo. 2024. From" um" to" yeah": Producing, predicting, and regulating information flow in human conversation. *arXiv preprint arXiv:2403.08890*.

Holly P Branigan, Martin J Pickering, Jamie Pearson, and Janet F McLean. 2010. Linguistic alignment between people and computers. *Journal of pragmatics*, 42(9):2355–2368.

Xusen Cheng and Lulu Zhang. 2025. Inspiration booster or creative fixation? the dual mechanisms of llms in shaping individual creativity in tasks of different complexity. *Humanities and Social Sciences Communications*, 12(1):1–10.

Elizabeth Clark and Noah A. Smith. 2021. Choose your own adventure: Paired suggestions in collaborative writing for evaluating story generation models. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3566–3575, Online. Association for Computational Linguistics.

Umberto Domanti, Lorenzo Campidelli, Sergio Agnoli, and Antonella De Angeli. 2026. Are semantic networks associated with idea originality in artificial creativity? a comparison with human agents. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26, page 1–18. ACM.

Guillaume Dubuisson Duplessis, Caroline Langlet, Chloé Clavel, and Frédéric Landragin. 2021. Towards alignment strategies in human-agent interactions based on measures of lexical repetitions. *Language Resources and Evaluation*, 55(2):353–388.

Pascale Feldkamp, Jan Kostkan, Ea Overgaard, Mia Jacobsen, and Yuri Bizzoni. 2024. Comparing tools for sentiment analysis of danish literature from hymns to fairy tales: Low-resource language and domain challenges. In *Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 186–199.

Halfdan Nordahl Fundal and Yuri Bizzoni. 2026. Directional alignment and narrative agency in human-llm co-writing. *Preprint*, arXiv:2604.23676.

Elaine Hatfield, John T Cacioppo, and Richard L Rapson. 1993. Emotional contagion. *Current directions in psychological science*, 2(3):96–100.

Qiyue Hu, Bin Liu, Mads Rosendahl Thomsen, Jianbo Gao, and Kristoffer L Nielbo. 2021. Dynamic evolution of sentiments in never let me go: Insights from multifractal theory and its implications for literary analysis. *Digital Scholarship in the Humanities*, 36(2):322–332.

Theodora Koulouri, Stanislao Lauria, and Robert D Macredie. 2016. Do (and say) as i say: Linguistic adaptation in human–computer dialogs. *Human–Computer Interaction*, 31(1):59–95.

Qixin Li, Haocheng Lu, and Gaowu Wang. 2025. When ai speaks, do we follow? phonetic entrainment in human-ai dialogues. In *National Conference on Man-Machine Speech Communication*, pages 167–186. Springer.

Simone Luchini, Nina Lauharatanahirun, and Roger Beaty. 2025. Creative or uncreative partner: Comparing humans and ai in collaborative creative tasks.

Laurits Lyngbaek, Pascale Feldkamp, Yuri Bizzoni, Kristoffer Nielbo, and Kenneth Enevoldsen. 2026. Is sentiment banana-shaped? exploring the geometry and portability of sentiment concept vectors. In *The Proceedings for the 15th Workshop on Computational Approaches to Subjectivity, Sentiment Social Media Analysis (WASSA 2026)*, page 146–160. Association for Computational Linguistics.

Eliot Maës, Philippe Blache, and Leonor Becerra. 2022. Shared knowledge in natural conversations: can entropy metrics shed light on information transfers? In *Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL)*, pages 213–227, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Saif M Mohammad. 2016. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier.

Shakked Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192.

Emily Öhman and Riikka H Rossi. 2022. Computational exploration of the origin of mood in literary texts. In *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*, pages 8–14.

Rachel Ostrand, Victor S Ferreira, and David Piorkowski. 2023. Rapid lexical alignment to a conversational agent. In *INTERSPEECH*, pages 2653–2657.

Martin Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *The Behavioral and brain sciences*, 27:169–90; discussion 190.

Soujanya Poria, Navonil Majumder, Rada Mihalcea, and Eduard Hovy. 2019. Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE access*, 7:100943–100953.

Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ data science*, 5(1):31.

Min Tang, Sebastian Hofreiter, Christian H Werner, Aleksandra Zielińska, and Maciej Karwowski. 2025. "who" is the best creative thinking partner? an experimental investigation of human–human, human–internet, and human–ai co-creation. *The Journal of Creative Behavior*, 59(3):e1519.

Giovanna Varni, Isabelle Hupont, Chloe Clavel, and Mohamed Chetouani. 2017. Computational study of primitive emotional contagion in dyadic interactions. *IEEE Transactions on Affective Computing*, 11(2):258–271.

Emily Wenger and Yoed Kenett. 2025. We're different, we're the same: Creative homogeneity across llms. *Preprint*, arXiv:2501.19361.

Yang Xu and David Reitter. 2018. Information density converges in dialogue: Towards an information-theoretic model. *Cognition*, 170:147–163.

Eric B Zhou, Dokyun Lee, and Bin Gu. 2025. Who expands the human creative frontier with generative ai: Hive minds or masterminds? *Science Advances*, 11(36):eadu5800.

A Appendix

A.1 Recruitment and Preprocessing

The Human–LLM condition is drawn from Fundal and Bizzoni (2026). For the Human–Human condition, participants were recruited from English-taught university and graduate programs at Aarhus University; participation was voluntary and uncompensated, and participants were paired anonymously in the writing interface. The LLM–LLM condition was generated with the same prompt and turn-taking structure, using paired instances of the models used in the Human–LLM condition.

Before analysis, we removed incomplete sessions, empty turns, and malformed model outputs. Because all alignment and influence measures depend on adjacent turns, stories with missing turns contributed only valid adjacent exchange pairs. Human-written turns were spell-corrected before metric extraction, after which valence, embedding, semantic-distance, and surprisal-based measures were computed on the final analytic dataset. We used LLM spell-correction for human contributions.

Additional diagnostics are shown in Figures 6, 7, 8, 9, 10, 11, 12. Full metric-level asymmetry estimates are reported in Table 5.

Metric	Condition	n	Mean A_m	95% CI	Cliff's δ vs. HA
Novelty	Human–LLM	97	0.828	[0.732, 0.926]	–
Novelty	Human–Human	36	0.355	[0.284, 0.428]	0.591
Novelty	LLM–LLM	80	0.209	[0.175, 0.245]	0.766
Transience	Human–LLM	97	0.484	[0.432, 0.537]	–
Transience	Human–Human	36	0.300	[0.233, 0.370]	0.397
Transience	LLM–LLM	80	0.197	[0.165, 0.231]	0.670
Resonance	Human–LLM	97	1.290	[1.149, 1.432]	–
Resonance	Human–Human	36	0.605	[0.482, 0.733]	0.572
Resonance	LLM–LLM	80	0.313	[0.254, 0.382]	0.806

Table 5: Metric-level slot asymmetry across conditions. A_m denotes the per-story absolute difference between slot-level mean metric values. Human–LLM shows substantially larger side-level asymmetry in novelty, transience, and resonance than both comparison conditions. Cliff's δ values compare Human–LLM against each baseline condition.

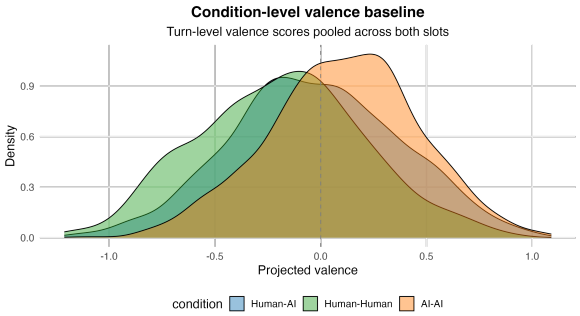

Figure 6: Distribution of turn-level valence scores by condition.

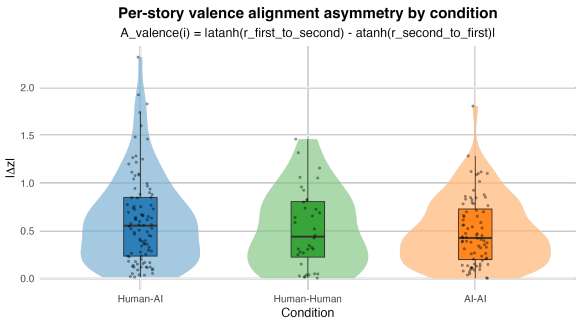

Figure 7: Per-story valence asymmetry by condition.

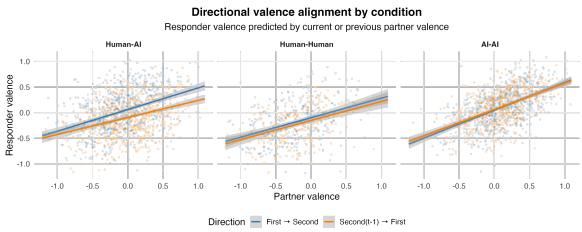

Figure 8: Directional valence alignment diagnostics across HH, HA, and AA conditions.

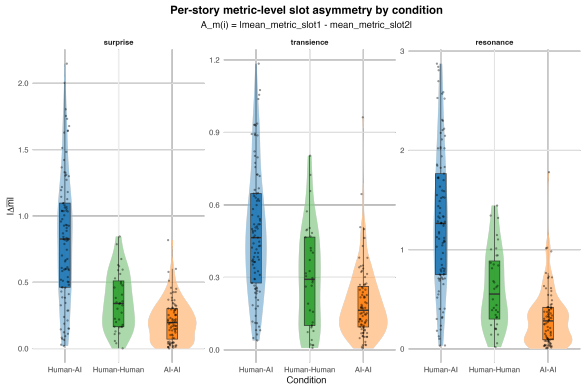

Figure 9: Per-story metric-level slot asymmetry A_m by condition for Novelty, Transience, and Resonance.

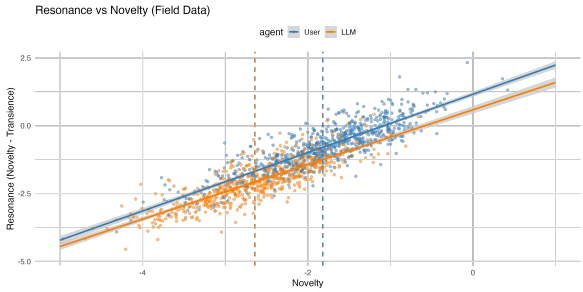

Figure 10: Novelty–resonance relationship within the HA condition.

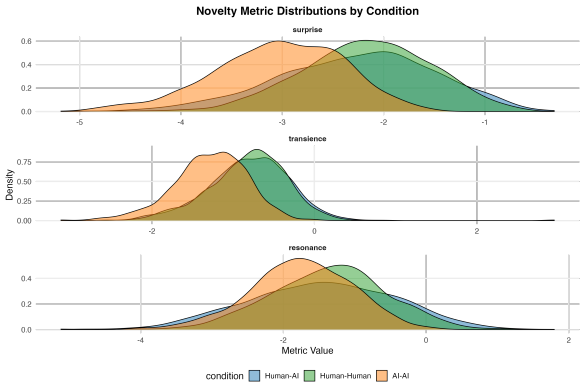

Figure 11: Novelty distributions by condition.

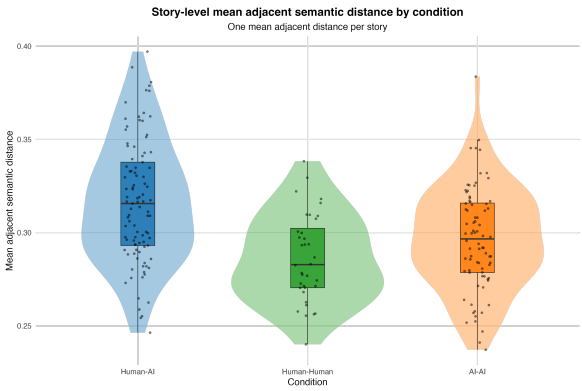

Figure 12: Story-level semantic distance by condition.