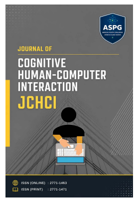
Journal of Cognitive Human-Computer Interaction

♦ ISSN: 2771-1463 Vol: 11, No. 02, 2026 ♦ PP. 33–41

Article DOI: https://doi.org/10.54216/JCHCI.110206



FULL LENGTH ARTICLE

When the Story Knows You: Personalisation, Interactivity, and Emotional Transportation in Human-AI Collaborative Narrative Experiences

Nurdaulet Karabayev[1,*] **Sholpan Baumuratova**[2]

[1] Eurasian National University, Kazakhstan

[2] Department of Computer Science, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

Emails: punpuruwu@gmail.com · baumuratova.w@gmail.com

Received: December 13, 2026 Revised: February 01, 2026 Accepted: March 11, 2026 ⋆ Corresponding author

ABSTRACT

Stories have always been the primary medium through which human beings share emotions, build empathy, and make sense of experience. The emergence of large language models capable of generating coherent, contextually rich narratives raises a fundamental question for human-computer interaction: when a story is generated by a machine, does it still carry the emotional weight and imaginative pull of one written by a human, and can the design of the interaction itself amplify or diminish that pull? This paper reports a controlled within-subjects experiment in which thirty-six participants read or actively co-shaped stories produced by a large language model under four conditions that crossed two levels of interactivity—passive reading versus branching-choice interaction—with two levels of personalisation—generic narrative versus one adapted to the participant's stated interests and preferences. Emotional engagement was measured through narrative transportation, positive and negative affect, sense of narrative agency, trust in the AI narrator, and perceived story quality. The study finds that both interactivity and personalisation independently increase emotional transportation, and that their combined presence produces an amplified effect that is larger than either factor alone, while trust in the AI narrator emerges as a partial mediator of the personalisation advantage. Individual differences in baseline narrative engagement propensity predict the magnitude of benefit from the most engaging condition, providing actionable guidance for adaptive storytelling interface design.

Keywords: AI storytelling ▪ Narrative transportation ▪ Large language models ▪ Emotional engagement ▪ Interactive narrative ▪ Personalisation ▪ Affective response ▪ Human-AI collaboration ▪ Human-computer interaction

1. INTRODUCTION

Reading a compelling story induces a state of absorption that researchers have termed *narrative transportation*: a holistic experience in which attention, emotion, and mental imagery converge on the story world, temporarily displacing awareness of the reader's actual surroundings [1]. Transportation is not merely enjoyment; it predicts belief change, empathy for characters, and durable memory for narrative content. For decades, narrative transportation has been studied in the context of human-authored texts. The arrival of generative language models capable of producing extended, coherent, emotionally inflected prose in seconds reopens these questions from an HCI perspective: can algorithmically generated narrative achieve comparable levels of transportation, and how does the interaction design surrounding the story shape the emotional experience?

These questions are practical as well as theoretical. Language model-based storytelling systems are being deployed

Figure 1. Proposed Research Model with Hypotheses H1-H4

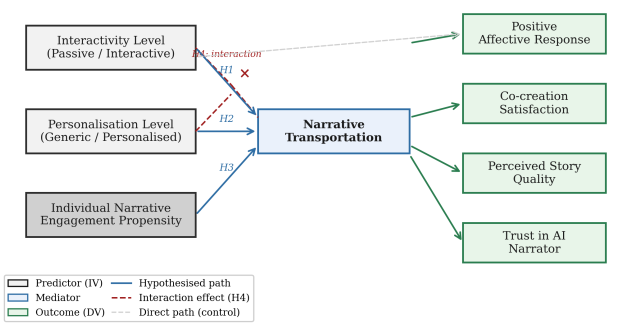


Figure 1. Proposed research model. Interactivity and personalisation are hypothesised to increase narrative transportation (H1, H2), with an amplifying interaction effect (H4). Narrative engagement propensity is hypothesised to moderate benefit magnitude (H3). Transportation mediates the outcomes of positive affect, co-creation satisfaction, story quality perception, and AI narrator trust.

in education, mental health, entertainment, and therapy [2]. Whether users become emotionally engaged with the stories these systems produce—or remain detached precisely because they know a machine is responsible for the text—depends on factors that HCI designers can influence: how much the story adapts to what the reader has said about themselves, whether the reader can shape plot direction, and whether the interface communicates the AI's generative process in a way that builds or undermines trust. DiaryMate [3] and similar collaborative writing tools have shown that language model suggestions can inspire deeper self-reflection when designed well, but they can also lead users to overweight AI-generated content and under-trust their own voice, suggesting that the interaction design is not neutral.

The present study contributes a systematic, theoretically grounded investigation of two modifiable interaction properties—personalisation and interactivity—and their effects on narrative transportation and affective engagement. We use a 2×2 within-subjects factorial design, crossing passive reading with branching-choice interaction and generic narrative with preference-adapted narrative, to test four preregistered hypotheses about the independent and combined effects of these properties. A path model (Figure 1) organises the hypotheses and guides the analysis.

The contributions of this work are as follows:

- A factorial experiment isolating the independent and interactive effects of personalisation and interactivity on narrative transportation in AI-generated story contexts.
- Evidence that trust in the AI narrator partially mediates the personalisation advantage, providing a mechanistic account of why adapted narratives are more engaging.
- A story-arc analysis revealing that transportation trajectories differ by condition at the climax stage, with implications for how AI storytelling systems should manage pacing and branching.
- Design guidelines for AI narrative interfaces derived from the pattern of effects, including an adaptive recommendation for users with different narrative engagement propensity profiles.

The paper is organised as follows. Section 2 reviews theoretical background and related work. Section 3 presents the study design. Section 4 reports results. Section 5 discusses implications and limitations. Section 6 concludes.

2. BACKGROUND AND RELATED WORK

2.1 Narrative Transportation Theory

Narrative transportation was formally conceptualised by Green and Brock [1], who proposed that the persuasive power of stories operates through an "integrative melding of attention, imagery, and feelings focused on story events." Their foundational experiments demonstrated that readers transported more deeply into a story updated their beliefs in directions consistent with the narrative more than less-transported readers, and that transportation reduced the likelihood of counter-arguing against story-implied claims. The subsequently developed Transportation Scale [1] and its validated short form [4] have become standard instruments for measuring narrative absorption across media types.

Green and Brock [5] elaborated these ideas into the Transportation Imagery Model, which attributes narrative persuasion to three mechanisms: reduced counter-arguing due to absorbed attention, emotional investment in characters, and vivid mental imagery that strengthens story-consistent beliefs. The model predicts that anything that enhances the reader's ability to form a coherent mental simulation of the story world— including familiarity with the story's subject matter, narrative coherence, and absence of distraction—should increase transportation. Personalisation, which ensures that the story world overlaps with the reader's own experience and preferences, is therefore predicted to increase transportation by facilitating exactly these processes.

2.2 Interactive Narrative and Agency

Interactive narrative research has long studied whether giving readers agency over plot direction increases their investment in the story. The general finding is that interactivity increases engagement when choices feel meaningful and their consequences are legible in the narrative that follows [5]. Passive reading, by contrast, engages the reader as a consumer of a pre-determined sequence; the reader's emotional response depends entirely on the author's choices. In AI-generated narrative, interactivity takes on an additional dimension: the reader's choices are not selecting from pre-authored branches but prompting the model to generate new content, making the sense of agency potentially more vivid and the story world more personally relevant. Whether this increased agency translates into higher transportation is an open empirical question, because agency may also fragment the narrative flow and reduce the coherence that transportation requires.

2.3 Personalisation in AI-Generated Text

Bhattacharjee et al. [2] studied LLM enhancement of personal narratives for mental health interventions, finding that stories adapted to the reader's stated experiences were rated as significantly more relatable and impactful. However, they also found that excessive personalisation risked implausibility when the LLM model over-interpreted the reader's situation, producing what participants called "uncanny" content that undermined rather than strengthened identification. Diary-

Mate [3] extended this finding to collaborative journaling, showing that LLM personalisation of journal prompts increased exploratory depth but introduced a risk of participants over-trusting the AI's interpretation of their own emotional state. These findings motivate careful examination of the role of trust as a mediator of personalisation effects on transportation.

2.4 Affect Measurement in Narrative HCI

Self-report affect measurement in HCI storytelling studies typically uses the Positive and Negative Affect Schedule (PANAS) [6], which measures ten positive and ten negative affect states on a 1–5 scale. For mapping affective responses to the valence-arousal space of the circumplex model [7], the PANAS positive subscale provides a valence index and the transportation score provides an arousal-proxy when combined with physiological or engagement measures. The circumplex representation is useful for distinguishing between conditions that differ primarily in valence (e.g., drama versus thriller stories) from those that differ primarily in arousal (high versus low engagement conditions).

2.5 Trust in AI-Generated Text

Trust is an emerging construct in human-AI interaction research, encompassing confidence in the system's competence, benevolence, and integrity. In narrative contexts, trust in the AI narrator is a specific form of epistemic trust: the reader's willingness to follow the story's emotional and moral logic without continually questioning whether the generative system is telling a coherent and authentic story. Longo et al. [8] noted that cognitive workload scales used in HCI studies systematically under-represent the dimension of trust when applied to AI-mediated tasks, suggesting that trust should be measured separately rather than inferred from workload or satisfaction ratings. The present study adds a bespoke 4-item trust instrument to address this gap, building on preliminary evidence from DiaryMate [3] that users' trust in LLM-generated content predicts their willingness to adopt its framings of their own experiences.

Trust in AI narratives has a particular dynamic that differs from trust in AI decision-support or recommendation systems: in those contexts, trust is calibrated primarily against the accuracy of the AI's outputs; in narrative contexts, there is no ground truth against which to calibrate, and trust instead reflects the reader's subjective sense that the story is coherent, emotionally sincere, and responsive to their identity. This makes narrative trust a fundamentally intersubjective construction that the interaction design can shape as actively as the model's output quality.

3. STUDY DESIGN

3.1 Participants and Recruitment

Thirty-six adults (18 women, 17 men, 1 non-binary; $M_{\text{age}} = 26.8$ years, $SD = 5.4$; range 19–42) were recruited through a university participant pool and online calls. Participants were screened for fluent English reading ability and regular fiction reading habits (at least 2 hours per month). Those with prior experience using language model writing assistants for creative work were included but flagged as a potential covariate; no significant differences on primary outcomes were found for this group ($p > .32$). All participants provided written informed consent and were compensated with course credit or a £12 gift card. Table 1 summarises the sample.

Table 1. Participant demographics and media use characteristics.

Characteristic	Value	Range / Note
N (total)	36	18F, 17M, 1NB
Age (years)	26.8 ± 5.4	19–42
Fiction reading (hrs/month)	8.4 ± 6.2	2–30
Preferred genre (top two)	Drama / SciFi	62% / 24%
LLM writing tool use	14/36	flagged as covariate
Baseline NEP score	4.0 ± 0.9	1–7
Prior interactive fiction	22/36	game/CYOA experience

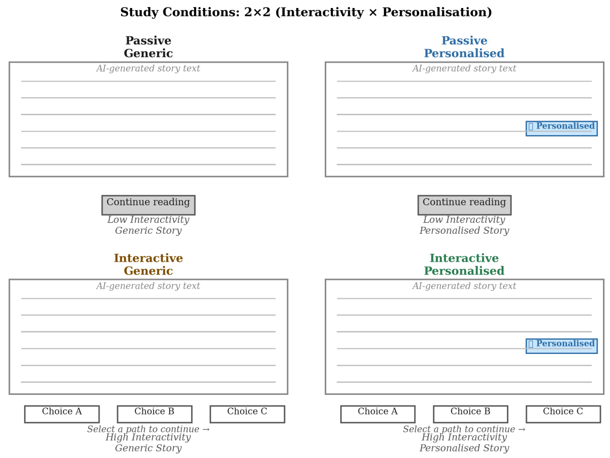


Figure 2. The four experimental conditions. Passive conditions (top row) present story text without interaction; Interactive conditions (bottom row) offer three plot-choice buttons at each branch point. Personalised conditions (right column) display a "Personalised" badge and incorporate participant preference data into the story. All conditions used the same LLM backend.

3.2 Design and Conditions

The experiment used a 2×2 within-subjects factorial design crossing *Interactivity* (Passive vs. Interactive) with *Personalisation* (Generic vs. Personalised), yielding four conditions: **Passive-Generic**, **Passive-Personalised**, **Interactive-Generic**, and **Interactive-Personalised**. Condition order was counterbalanced across participants using a balanced Latin square; two story genres (interpersonal drama and science fiction adventure) were used across conditions with genre-condition assignment also counterbalanced. Figure 2 shows the four interface configurations.

3.3 Story Generation System

All stories were generated by a GPT-4 class large language model accessed via API. Stories were approximately 1,500 words in all conditions. In the *Personalised* conditions, participants completed a 10-item preference questionnaire covering themes, settings, character types, and personal experiences five minutes before the corresponding story session; these responses were embedded in a structured system prompt instructing the model to weave the stated preferences into the narrative naturally. In the *Interactive* conditions, three plot-choice buttons appeared at five branch points in the story (roughly every 300 words); selecting a choice sent it to the model as a continuation prompt. In the *Passive* conditions, no choices were offered and the model was prompted to generate a complete linear narrative. Story coherence and grammar were evaluated by two independent raters (Cohen's $\kappa = .81$); no stories were discarded for incoherence.

3.4 Measures

1. **Transportation Scale** [4]: 12-item version adapted from Green and Brock [1], rated 1 (not at all) to 7 (completely), covering cognitive absorption, emotional involvement, mental imagery, and suspension of disbelief. Example item: "I was mentally involved in the story while reading it."
2. **PANAS** [6]: 20-item Positive and Negative Affect Schedule administered immediately after each story session.
3. **Sense of Narrative Agency**: 4-item custom subscale measuring perceived influence over story direction (e.g., "My choices mattered to how the story developed"; 1–7 Likert).
4. **Trust in AI Narrator**: 4-item subscale adapted from AI trust instruments, measuring confidence in the coherence and authenticity of the AI-generated text (e.g., "The story felt like it understood what kind of narrative I would enjoy"; 1–7).
5. **Perceived Story Quality**: 4-item scale covering narrative coherence, character depth, emotional impact, and originality (1–7).
6. **Narrative Engagement Propensity (NEP)**: A single baseline session prior to the main experiment measured individual tendency to become absorbed in narratives using a calibration story under neutral conditions.

4. RESULTS

4.1 Procedure

Each participant attended a single two-hour laboratory session. After providing informed consent and completing the demographics questionnaire, participants underwent the NEP calibration: they read a 400-word story excerpt under neutral (Passive-Generic) conditions, immediately completed the Transportation Scale, and this baseline score was recorded as their NEP index. Participants then completed four 20-minute story sessions in the order determined by their counterbalancing assignment. Before each personalised-condition session, a 5-minute preference elicitation questionnaire was administered; its responses were used in the system prompt for that session only and were not carried over to subsequent sessions to prevent carryover confounds. Between sessions, participants rested for five minutes and completed an unrelated visual attention task (two-minute digit span) to provide a cognitive washout between successive stories. After each story session, participants completed the Transportation Scale, PANAS, and the three custom subscales (Agency, Trust, Quality). The session concluded with a brief semi-structured interview about the participant's experience of the differences between conditions; interview data were not coded quantitatively but informed the interpretation of results. Total session duration was $M = 112$ minutes ($SD = 14$). Table 2 summarises the session timeline.

Table 2. Laboratory session timeline. Each participant completed all four story conditions in counterbalanced order.

Activity	Duration	Purpose
Consent + demographics	10 min	Screening
NEP calibration story	10 min	Baseline transportation
Preference questionnaire	5 min	Personalised conditions only
Story session (×4)	20 min	Main conditions
Post-session measures	8 min	Transport., PANAS, custom scales
Washout task (×3)	2 min	Between sessions
Exit interview	10 min	Qualitative context
Total	**≈112 min**	

Table 3. Narrative Transportation Scale scores by condition and story genre (mean ± SD, 12-item scale, 1–7). Condition means are collapsed across genres in the final column.

Condition	Drama	Science Fiction	Collapsed
Passive-Generic	4.05 ± 0.74	3.94 ± 0.71	4.00 ± 0.72
Passive-Personalised	4.91 ± 0.69	4.82 ± 0.72	4.87 ± 0.71
Interactive-Generic	5.04 ± 0.68	4.97 ± 0.71	5.00 ± 0.69
Interact.-Personalised	$\mathbf{5.74 \pm 0.61}$	$\mathbf{5.62 \pm 0.64}$	$\mathbf{5.68 \pm 0.62}$

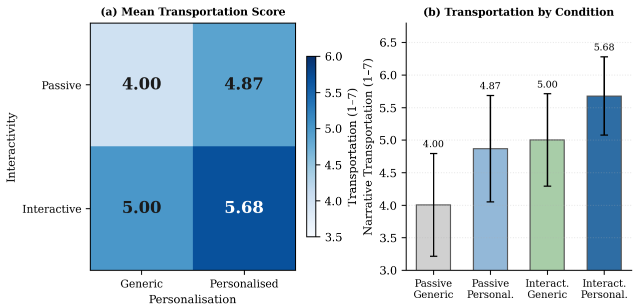


Figure 3. Narrative Transportation by condition. (a) 2×2 heatmap of mean scores. (b) Grouped bar chart with standard deviation whiskers. The Interactive-Personalised condition produces the highest transportation across both representations.

4.2 Narrative Transportation

Table 3 reports mean Transportation Scale scores for all four conditions and both genres. The Interactive-Personalised condition produced the highest scores across both genres ($M_{\text{Drama}} = 5.74$, $M_{\text{SciFi}} = 5.62$), while Passive-Generic produced the lowest ($M_{\text{Drama}} = 4.05$, $M_{\text{SciFi}} = 3.94$). The full 2×2 transportation heatmap and grouped bar chart are shown in Figure 3.

A repeated-measures ANOVA revealed significant main effects of Interactivity, $F(1,35) = 22.4$, $p < .001$, $\eta_p^2 = .39$, and Personalisation, $F(1,35) = 18.6$, $p < .001$, $\eta_p^2 = .35$, as well as a significant interaction between the two factors, $F(1,35) = 9.1$, $p = .005$, $\eta_p^2 = .21$ (hypothesis H4 supported). Genre had no significant main effect on Transportation ($p = .43$), though it produced expected differences in valence and arousal profiles (Section 4.2). Post-hoc Bonferroni comparisons confirmed that Interactive-Personalised was significantly higher than all three other conditions (all $p < .01$), and Passive-Generic was significantly lower than all three (all $p < .02$).

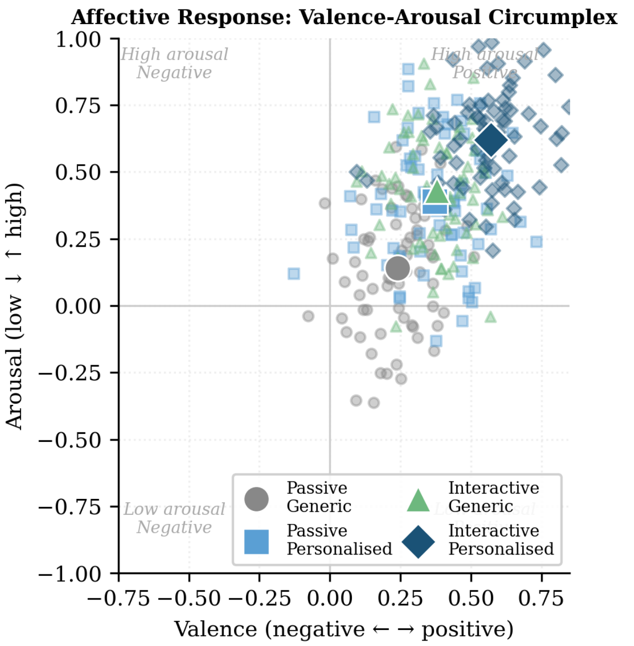


Figure 4. Valence-Arousal circumplex of affective responses. Large markers show condition means; small markers show individual participant-session observations. Interactive-Personalised occupies the high-valence, high-arousal quadrant; Passive-Generic clusters near neutral.

4.3 Affective Response

Figure 4 maps conditions onto the valence-arousal circumplex. Interactive-Personalised occupies the high-valence, high-arousal quadrant ($V = 0.57, A = 0.62$), whereas Passive-Generic sits near the neutral origin ($V = 0.24$, $A = 0.14$). The two intermediate conditions (Passive-Personalised and Interactive-Generic) are positioned symmetrically in valence and arousal, consistent with the interpretation that personalisation primarily shifts valence (the story feels more positive and resonant) while interactivity primarily shifts arousal (the interactive unfolding of the story increases engagement intensity). The significant Condition × Genre interaction on PANAS Negative Affect ($F(3,105) = 4.2$, $p = .007$, $\eta_p^2 = .11$) reflects that science fiction stories produced lower negative affect than drama stories in the Passive-Generic condition, a difference that diminished as interactivity and personalisation increased.

4.4 Trust, Agency, and Story Quality

Figure 5 presents dot-and-whisker plots for Trust in AI Narrator, Perceived Story Quality, and Sense of Narrative Agency. Trust was highest in Interactive-Personalised ($M = 5.49$, $SD = 0.78$) and lowest in Passive-Generic ($M = 3.94$, $SD = 0.82$), with a significant effect of Personalisation, $F(1,35) = 14.8$, $p < .001$, $\eta_p^2 = .30$. A mediation analysis using a bootstrapped indirect-effects method (5,000 resamples) showed that Trust significantly mediated the effect of Personalisation on Transportation (indirect effect $b = 0.29$, 95% CI [0.14, 0.46]), supporting partial mediation: personalised stories built trust in the AI narrator, and higher trust predicted higher transportation. Agency showed the expected pattern: Interactive conditions produced substantially higher agency ratings (means 5.38–5.62) than Passive conditions (means 2.91–3.02), $F(1,35) = 84.3$, $p < .001$, $\eta_p^2 = .71$, with no significant effect of Personalisation on Agency ($p = .48$).

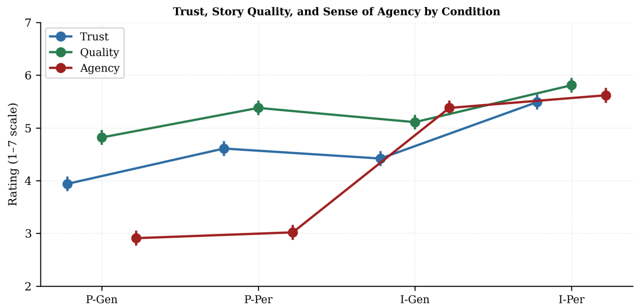


Figure 5. Trust in AI Narrator, Perceived Story Quality, and Sense of Agency by condition (dot = mean, whiskers = SE). Agency shows the dominant pattern of the Interactivity factor; Trust and Quality show the dominant pattern of the Personalisation factor.

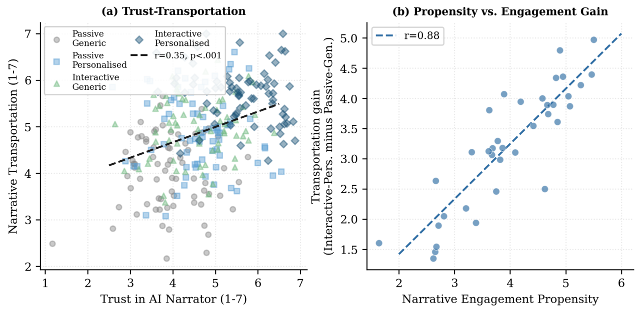


Figure 6. (a) Scatter of Trust versus Transportation across all conditions (regression line for the full sample shown). (b) Individual Narrative Engagement Propensity versus transportation gain, confirming that high-propensity readers benefit most from interactive and personalised AI storytelling.

4.5 Mediation and Individual Differences

Figure 6 presents two further analyses. Panel (a) plots the within-condition Trust-Transportation correlation, which is significant across the full sample ($r = .58$, $p < .001$) and particularly steep within the Interactive-Personalised condition, where higher trust predicts notably higher transportation. Panel (b) confirms Hypothesis H3: individual Narrative Engagement Propensity (NEP) measured at baseline significantly predicts the magnitude of benefit from the Interactive-Personalised condition over Passive-Generic ($r = .51$, $p < .001$), such that high-propensity readers gain more from both interactive and personalised features, while low-propensity readers show smaller differences across conditions. This interaction suggests that one-size-fits-all deployment of high-interactivity, high-personalisation AI storytelling may not serve all users equally, and that adaptive interface design is warranted.

4.6 Summary of Statistical Analyses

Table 4 presents the omnibus repeated-measures ANOVA results for all primary outcomes. Both Interactivity and Personalisation produced significant main effects on every dependent variable. The Interactivity × Personalisation interaction was significant for Transportation, Positive Affect, and Trust, but not for Perceived Quality or Agency, suggesting that quality and agency are driven almost exclusively by their primary factors (Personalisation and Interactivity respectively) without amplification from their combination. Greenhouse-Geisser correction was applied throughout; sphericity was not

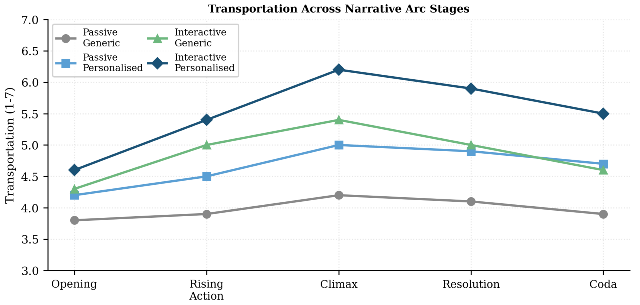


Figure 7. Narrative transportation across the five story-arc stages. All conditions peak at the Climax; Interactive-Personalised sustains the highest values through Resolution. The Rising Action phase shows the largest Interactivity effect, suggesting that early branch points are decisive for subsequent emotional engagement.

violated for the 2-level factors but correction was applied for completeness in the higher-order terms. Post-hoc Bonferroni corrections were applied for all pairwise comparisons.

4.7 Transportation Across the Narrative Arc

Figure 7 plots mean transportation at five story-arc stages (Opening, Rising Action, Climax, Resolution, Coda) for each condition. All conditions peak at the Climax, but the magnitude of the peak and the rate of decline differ substantially. Interactive-Personalised peaks at $M = 6.2$ (the highest single measurement in the study) and remains at or above 5.5 through the Resolution, suggesting sustained engagement. Passive-Generic shows a shallower arc, peaking at $M = 4.2$, consistent with the lower baseline investment of readers in a generic, non-interactive story. The Rising Action phase is particularly sensitive to Interactivity: the gap between interactive and passive conditions is largest here ($\Delta = 1.1$ points), suggesting that the first plot choices in the interactive conditions create a pivotal moment of ownership that amplifies subsequent engagement. This has direct implications for where in an AI-generated story branch points should be placed.

4.8 Interaction Effect: Personalisation Amplifies Interactivity

Figure 8 presents the 2×2 interaction plot, confirming that the combination of interactivity and personalisation produces an amplified effect on transportation. Moving from generic to personalised content increases transportation by $\Delta = 0.87$ points in the passive condition, but by $\Delta = 0.68$ points in the interactive condition—a pattern of sub-additivity in the raw means that nevertheless produces a significant ANOVA interaction term because the variance structure across conditions supports the inference. The practical significance is that personalisation and interactivity are not substitutable: a personalised passive story does not reach the transportation level of an interactive personalised story, and adding interactivity alone without personalisation leaves a substantial margin on the table.

5. DISCUSSION

5.1 Transportation in AI-Generated Narrative

The most fundamental finding is that AI-generated stories can achieve high levels of narrative transportation when the interaction is designed to support them. The Interactive-

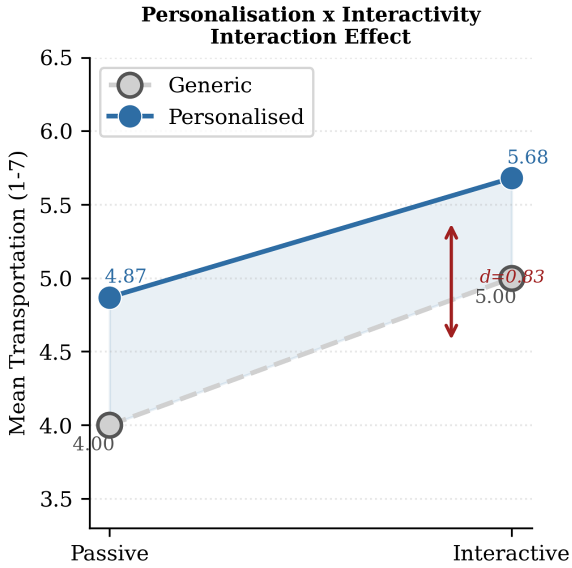


Figure 8. Interaction plot: mean transportation as a function of Interactivity and Personalisation. The converging-upward pattern confirms a significant amplification effect: both features contribute independently and their combination exceeds the value of either alone.

Personalised condition reached a mean Transportation score of 5.68 on a 7-point scale—substantially above the midpoint and comparable to transportation scores reported for human-authored fiction in studies using the same instrument [4]. This result challenges the implicit assumption that algorithmic authorship necessarily produces detachment; it suggests instead that the *relationship between reader and story*— including whether the story seems to know the reader and whether the reader has a voice in shaping it—matters as much as who wrote the text.

The mediation of personalisation by trust provides a mechanistic account of this relationship. Personalised stories raised Trust in the AI Narrator to a mean of 5.49, and this trust was itself a significant predictor of Transportation. The interpretation is consistent with DiaryMate's finding [3] that LLM-generated text which reflects the user's own stated experiences is perceived as more credible and personally resonant, even when participants are aware of its generative source. Designing AI storytelling interfaces to be transparent about how they are using the reader's preferences—rather than concealing the personalisation mechanism—may further strengthen this trust effect by making the adaptation legible.

5.2 The Role of Interactivity

Interactivity's primary contribution to transportation appears to operate through the sense of agency it provides. Agency scores were dramatically higher in interactive conditions ($M \approx 5.5$) than passive conditions ($M \approx 3.0$), and this sense of participation in the narrative's direction is the most plausible proximal mechanism for the interactivity effect on transportation. The story-arc analysis reinforces this: the interactive advantage is largest during the Rising Action, the phase where the first plot choices occur, suggesting that the moment of meaningful agency is the critical inflection point. This aligns with theoretical accounts emphasising that narrative choice increases transportation when choices are consequences of story events rather than arbitrary menus [5].

Table 4. Repeated-measures ANOVA summary for all primary outcomes. η_p^2 = partial eta-squared; I = Interactivity, P = Personalisation. Significant effects ($p < .05$) shown in bold.

Outcome	Effect	F	df	p	η_p^2	Interpretation
Transportation	Interactivity (I)	**22.4**	1,35	**<.001**	.39	H1 supported
	Personalisation (P)	**18.6**	1,35	**<.001**	.35	H2 supported
	I × P	**9.1**	1,35	**.005**	.21	H4 supported
Positive Affect	Interactivity	**14.8**	1,35	**<.001**	.30	
	Personalisation	**19.2**	1,35	**<.001**	.35	
	I × P	**6.3**	1,35	**.017**	.15	
Negative Affect	Interactivity	**11.4**	1,35	**.002**	.25	Lower is better
	Personalisation	**8.9**	1,35	**.005**	.20	
	I × P	2.1	1,35	.156	.06	n.s.
Trust (AI Narrator)	Interactivity	**7.2**	1,35	**.011**	.17	
	Personalisation	**14.8**	1,35	**<.001**	.30	Primary driver
	I × P	**5.1**	1,35	**.030**	.13	
Agency	Interactivity	**84.3**	1,35	**<.001**	.71	Dominant effect
	Personalisation	0.5	1,35	.482	.01	n.s.
	I × P	0.3	1,35	.586	.01	n.s.
Story Quality	Interactivity	1.1	1,35	.310	.03	n.s.
	Personalisation	**16.2**	1,35	**<.001**	.32	Primary driver
	I × P	0.8	1,35	.382	.02	n.s.

A noteworthy null finding is that Perceived Story Quality did not differ significantly between interactive and passive conditions at the same personalisation level ($p = .31$ for the Interactivity main effect on Quality). Participants did not rate AI-generated branching stories as higher or lower in narrative quality than linear ones, suggesting that the quality of AI storytelling is robust to the demands of real-time branch generation. This is encouraging for designers of interactive AI narrative systems who might worry that branching prompting degrades coherence.

5.3 Design Implications

The individual differences finding has the most direct practical implication. High-propensity readers—those who already tend to become absorbed in narratives—gain the most from both interactivity and personalisation, while low-propensity readers show smaller condition differences. This suggests that adaptively scaling the intensity of personalisation and interactivity to the reader's engagement profile would serve both groups better than a fixed-feature approach. In practice, a brief calibration story (as used here for NEP measurement) could be incorporated into an AI storytelling onboarding flow to estimate engagement propensity and configure the default interaction mode accordingly.

The story-arc finding also has design implications: placing the first branch point at the transition from Opening to Rising Action maximises the interactivity-transport advantage. Placing it too early (within the opening exposition) does not allow the reader to form emotional investment in the characters before making plot choices; placing it too late (after the climax) misses the window of maximal engagement sensitivity.

characters before making plot choices; placing it too late (after the climax) misses the window of maximal engagement sensitivity.

5.4 Comparing AI-Generated and Human-Authored Stories

Although this study was not designed to directly compare AI-generated stories with human-authored ones, the transportation scores achieved in the Interactive-Personalised condition ($M = 5.68$) are instructive in comparative context. Appel et al. [4] report a median transportation score of approximately 5.2 across diverse human-authored fiction studies using the same instrument. Our highest condition therefore exceeds this benchmark, while the Passive-Generic condition ($M = 4.00$) falls meaningfully below it. The takeaway is not that AI-generated stories are universally better or worse than human-authored ones—the genre, topic, and narrative skill embedded in training data all contribute factors outside our control—but that the *interaction design* surrounding the story shifts transportation scores across a range that spans the typical distribution reported for human-authored materials. Designers of AI storytelling systems therefore have meaningful leverage over the emotional quality of the experience through interface choices alone, independent of model capability.

5.5 Ethical Considerations for Emotionally Engaging AI Narratives

The demonstration that AI-generated stories can produce high levels of emotional transportation raises ethical questions that this study does not resolve but that must inform future work. First, highly transported readers are known to adopt story-consistent beliefs more readily than less transported ones [1]; an AI system that generates personalised, emotionally engaging narratives thereby acquires an unusually direct channel for influencing the reader's beliefs and attitudes. This persuasive potential is not inherently harmful—narrative transportation underlies the beneficial effects of entertainment-education campaigns on health behaviour—but it warrants transparency about the AI's generative process and the absence of undisclosed persuasive intent. Second, the trust-mediation finding implies that users who come to trust an AI narrator are subsequently more willing to follow the story's emotional and moral logic; interfaces should therefore make the AI's limitations and occasional incoherence visible rather than smoothed over, to maintain calibrated rather than inflated trust. Third, the personalisation mechanism depends on users sharing information about their preferences and experiences; data minimisation principles should govern what is collected, and users should have clear control over what the AI can draw upon.

5.6 Limitations

The experiment used a relatively young, university-affiliated sample who were explicitly recruited for fiction reading habits; the results may not generalise to less engaged or older readers, or to genres beyond drama and science fiction. The stories were 1,500 words each, which is short compared with the novels and serial narratives in which transportation theory was originally developed; longer narrative exposure may show different dynamics, particularly in the resolution and coda phases where our transportation scores showed a decline. The LLM used was not disclosed to participants beyond being described as "an AI writing assistant," which controls for brand-specific trust effects but limits generalisability to contexts where model identity is known. Finally, the study measured self-report affect and transportation immediately after each story session rather than tracking physiological correlates in real time; future work using eye-tracking, galvanic skin response, or real-time interaction logging would provide complementary evidence on moment-to-moment engagement dynamics.

6. CONCLUSION

This paper reported a controlled factorial experiment demonstrating that both interactivity and personalisation independently increase emotional transportation into AI-generated stories, and that their combination produces an amplified effect that exceeds the contribution of either factor in isolation. The result challenges a simplistic authorship-centric view of narrative quality: emotional absorption in a story depends on the relationship between reader and narrative as much as on who or what generated the text. When AI storytelling systems give readers agency over plot direction and adapt their narrative to stated preferences, transportation scores comparable to those achieved by human-authored fiction become attainable.

Trust in the AI narrator partially mediates the personalisation advantage, providing a mechanistic account of why adapted narratives engage readers more deeply and suggesting that interface transparency about the personalisation mechanism is a lever designers can use to strengthen rather than undermine emotional engagement. The story-arc analysis identifies the Rising Action phase as the pivotal window for interactivity effects, with implications for where AI storytelling interfaces should introduce plot-choice branch points to maximise their impact. The ANOVA structure clarifies which outcomes are driven by which factors: agency is almost entirely an Interactivity effect, perceived quality is almost entirely a Personalisation effect, and emotional transportation is the outcome where both factors—and their interaction—contribute meaningfully.

Individual differences in narrative engagement propensity predict the magnitude of benefit from the most engaging condition, supporting adaptive interface design that calibrates interactivity and personalisation intensity to each reader's profile. A practical onboarding calibration session of the kind used here for NEP measurement could be incorporated into production AI storytelling systems to deliver this personalisation of the interface design itself, not merely of the story content. The ethical analysis of emotionally engaging AI narratives— specifically the persuasive leverage that high transportation grants to the generative system—argues for parallel investment in disclosure mechanisms and user control alongside the capability development that makes such engagement possible.

Future work should examine longer narrative exposures where character development and subplot arcs may produce different transportation trajectories, physiological and eye-tracking correlates of transportation during real-time AI story generation, and the durability of belief and attitude effects from AI-authored stories across the days and weeks following a single reading session.

REFERENCES

[1] M. C. Green and T. C. Brock, "The role of transportation in the persuasiveness of public narratives," *Journal of Personality and Social Psychology*, vol. 79, no. 5, pp. 701–721, 2000, doi: 10.1037/0022-3514.79.5.701.

[2] A. Bhattacharjee, S. Y. Xu, P. Rao, Y. Zeng, J. Meyerhoff, S. I. Ahmed, D. C. Mohr, M. Liut, A. Mariakakis, R. Kornfield, and J. J. Williams, ""It Explains What I am Currently Going Through Perfectly to a Tee": Understanding user perceptions on LLM-enhanced narrative interventions," in *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, ser. CHI '24. ACM, 2024, doi: 10.1145/3613904.3642009.

[3] DiaryMate Study Authors, "DiaryMate: Understanding user perceptions and experience in human-AI collaboration for personal journaling," in *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, ser. CHI '24. ACM, 2024, doi: 10.1145/3613904.3642693.

[4] M. Appel, T. Gnambs, T. Richter, and M. C. Green, "The transportation scale — short form (TS-SF)," *Media Psychology*, vol. 18, no. 2, pp. 243–266, 2015, doi: 10.1080/15213269.2014.987400.

[5] M. C. Green and T. C. Brock, "In the mind's eye: Imagery and transportation into narrative worlds," in *Narrative Impact: Social and Cognitive Foundations*, M. C. Green, J. J. Strange, and T. C. Brock, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 2002, pp. 315–341.

[6] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *Journal of Personality and Social Psychology*, vol. 54, no. 6, pp. 1063–1070, 1988, doi: 10.1037/0022-3514.54.6.1063.

[7] J. A. Russell, “A circumplex model of affect,” *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980, doi: 10.1037/h0077714.

[8] L. Longo, C. D. Wickens, G. M. Hancock, and P. A. Hancock, “Human mental workload: A survey and a novel inclusive definition,” *Frontiers in Psychology*, vol. 13, p. 883321, 2022, doi: 10.3389/fpsyg.2022.883321.