

Evaluating Generative AI Models for Game Narrative Design: A Comparative Study of ChatGPT, Claude, and NovelAI

Abstract

This paper presents my comparative analysis of three generative AI models—ChatGPT (GPT-4), Claude (Anthropic), and NovelAI—for narrative design tasks in game development. Building on my midterm research examining AI implementation in major gaming studios like Ubisoft, Naughty Dog, and Rockstar Games, I evaluated each model's performance across five criteria: creativity, emotional depth, genre fit, clarity/flow, and ethical soundness. Using three distinct storytelling scenarios common in role-playing and adventure games, my research shows significant differences in how these models approach game narratives. Claude demonstrated superior emotional depth and worldbuilding similar to Naughty Dog's approach, ChatGPT excelled in structural clarity that would complement Ubisoft's efficiency-focused systems, and NovelAI showed limitations for implementation but potential value in conceptual development. My findings offer insights into how developers might leverage different AI models throughout the game narrative pipeline, from ideation to implementation, while considering ethical implications for player experience and creative workflows.

1. Introduction

The integration of artificial intelligence into creative processes has accelerated rapidly across the game development industry. When I began this research, I was particularly interested in how major studios like Ubisoft, Naughty Dog, and Rockstar Games have started exploring AI-assisted tools for asset generation, dialogue creation, and narrative design. As these technologies evolve, developers face critical decisions about which AI models best suit specific creative tasks, particularly in narrative design—where player engagement, emotional resonance, and coherent storytelling remain paramount.

Game narrative design presents unique challenges for AI systems. Unlike linear storytelling, game narratives must accommodate player agency, branching storylines, and interactive dialogue while maintaining consistent character voice, world-building elements, and emotional impact. These requirements raised several questions that I wanted to explore: Can current generative AI models produce game-ready narrative content? Which models excel at different aspects of narrative design? And how might these tools complement existing creative workflows rather than replace human writers?

This study builds directly upon my previous research examining how three major gaming studios—Ubisoft, Naughty Dog, and Rockstar Games—implement AI in their narrative pipelines. That research identified three distinct approaches to AI integration: Ubisoft's focus on automation and efficiency through tools like Ghostwriter AI, Naughty Dog's emphasis on emotional depth and realism, and Rockstar's implementation of emergent storytelling systems. These industry approaches provided the foundation for my current study, which seeks to determine which commercially available AI models might best support each studio's distinctive narrative priorities.

I addressed these questions through a systematic evaluation of three leading generative AI models—ChatGPT (GPT-4), Claude (Anthropic), and NovelAI—across three narrative scenarios central to game storytelling: an emotional NPC farewell, a moral dilemma with branching dialogue, and a mysterious quest introduction. By analyzing their outputs against criteria specifically relevant to game narrative design, my research aims to provide game developers, narrative designers, and researchers with actionable insights into the current capabilities and limitations of AI-assisted storytelling tools.

1.1 Research Questions and Objectives

My study directly extends my previous research on AI implementation in major gaming studios and is guided by the following research questions:

1. How do current generative AI models compare in their ability to produce narrative content that meets the specific requirements of game storytelling?
2. Which AI models might best support the diverse approaches to narrative design exemplified by major studios like Ubisoft (automation), Naughty Dog (emotional depth), and Rockstar Games (emergent storytelling)?
3. How might different AI models be optimally integrated into game narrative design workflows to complement human creativity rather than replace it?

To address these questions, I set the following objectives:

- Develop a systematic evaluation framework specific to game narrative design requirements
- Apply this framework to assess three distinct AI models across multiple narrative scenarios
- Identify patterns of strength and limitation for each model across different storytelling tasks
- Provide actionable recommendations for game developers on optimal model selection and implementation
- Consider ethical implications of AI-assisted narrative design for player experience and creative attribution

Building on previous research into AI applications in game development, my study contributes to understanding how different AI models might integrate into various stages of the narrative development pipeline. The findings have implications not only for technical implementation but also for creative collaboration between human writers and AI systems in producing compelling interactive narratives.

2. Background and Related Work

2.1 AI in Game Development

The use of artificial intelligence in game development has evolved dramatically in recent years. Historically, AI in games primarily referred to non-player character behavior and procedural content generation (Yannakakis & Togelius, 2018). However, the emergence of sophisticated language models has expanded AI's potential role in narrative creation processes that were traditionally the exclusive domain of human writers.

Major studios have begun integrating AI tools into their workflows. Ubisoft's development of Ghostwriter, an AI tool to assist with NPC dialogue generation, represents one of the first major implementations of generative AI in AAA game narrative pipelines (O'Donnell, 2022). Similarly, research from IEEE Transactions on Games (Zhang et al., 2022) has documented the growing trend of using AI for rapidly prototyping game dialogue and narrative branches.

2.2 Game Narrative Design Principles

Game narrative design differs significantly from traditional storytelling mediums. As Schell (2019) notes, game narratives must balance authored content with player agency, creating what Jenkins (2004) terms "emergent narratives." This poses unique challenges for AI systems, which must generate content that accommodates player choice while maintaining narrative coherence.

The interactive nature of game stories requires particular attention to branching dialogue structures, character consistency across multiple player interactions, and narrative flexibility. Industry standards in role-playing and adventure games often include key narrative elements such as emotionally resonant character moments, moral dilemmas with meaningful consequences, and intriguing quest introductions that establish both atmosphere and stakes (Heussner et al., 2015).

2.3 Building on Previous Research: Industry AI Implementation

This study builds directly upon my previous research examining how three major game studios—Ubisoft, Naughty Dog, and Rockstar Games—implement AI in their communication and narrative development strategies. In my midterm project, I discovered distinct approaches to AI integration that informed this current study:

Ubisoft's Automation Approach: From my research, I found that Ubisoft prioritizes AI for automation and scalability, particularly through their Ghostwriter AI system that assists writers by generating first drafts of NPC dialogue. This allows their narrative teams to focus on main storylines while AI handles more generic conversations. I noticed that Ubisoft's approach emphasizes efficiency, allowing rapid generation of content for their expansive open-world games, but sometimes sacrifices depth for breadth (O'Donnell, 2022).

Naughty Dog's Realism Focus: Unlike Ubisoft, Naughty Dog employs AI primarily to enhance emotional depth and realism. Based on my analysis of their development videos, their AI systems focus on believable character animations, dynamic enemy behaviors, and emotionally intelligent NPCs that respond to player choices. Their approach values handcrafted narrative experiences augmented by AI rather than generated by it.

Rockstar's Dynamic World Emphasis: I was particularly interested in how Rockstar Games utilizes AI to create emergent storytelling and living worlds, with systems that enable NPCs to follow realistic daily routines and respond organically to player actions. Their AI-driven procedural storytelling adapts side missions and world events based on player behavior, creating unique narrative experiences for each player.

These industry implementations raised important questions about AI's capabilities for narrative design that I wanted to address in this study. If major studios are already integrating AI into their narrative pipelines in these diverse ways, which currently available AI models might best support different aspects of narrative design? The distinct approaches of these studios—automation (Ubisoft), emotional depth (Naughty Dog), and emergent storytelling (Rockstar)—directly informed the evaluation criteria I applied in this study: creativity, emotional depth, genre fit, clarity/flow, and ethical soundness.

By investigating which AI models excel at different aspects of narrative design, I hoped to provide practical insights for studios seeking to implement AI in ways that align with their specific narrative priorities, whether those prioritize Ubisoft's efficiency, Naughty Dog's emotional resonance, or Rockstar's dynamic storytelling.

3. Methodology

For this study, I designed a systematic approach to evaluate the capabilities of three generative AI models for game narrative design tasks. I wanted my methodology to simulate realistic game development scenarios while enabling objective comparison across models.

3.1 Selection of AI Models and Prompt Engineering Approach

I selected three generative AI models based on their prominence and accessibility to game developers:

ChatGPT (GPT-4): Developed by OpenAI, GPT-4 represents one of the most widely used large language models with robust capabilities across various creative tasks. It has been adopted by numerous game studios for prototyping and ideation.

Claude: Developed by Anthropic, Claude has gained recognition for its nuanced understanding of context and ethical reasoning capabilities. It has been less extensively tested in game development contexts, making it valuable for comparative analysis.

NovelAI: A specialized AI system designed specifically for creative writing and storytelling, offering a contrast to general-purpose language models. Its focus on narrative generation made it a logical inclusion for game storytelling evaluation.

These models represent different approaches to AI-assisted creative writing, from general-purpose (ChatGPT) to ethically-aligned (Claude) to domain-specific (NovelAI).

3.1.1 Prompt Engineering Methodology

I spent considerable time designing my prompt engineering process to ensure fair comparison across models while optimizing for high-quality narrative outputs:

Consistent Detail Level: I made sure each prompt included the same level of contextual information and specific requirements. For example, the NPC farewell prompt provided a setting description and emotional tone guidance for all models.

Domain-Specific Framing: I explicitly framed prompts in game development terminology (e.g., "NPC," "quest opening," "branching dialogue") to orient the models toward game-specific outputs rather than general creative writing.

Implementation-Focused Instructions: Rather than asking for conceptual ideas about game writing, I requested concrete, implementable content (e.g., "Write a conversation" rather than "Describe how you would approach writing").

No Model-Specific Optimization: I used the same prompt text across all three models without tailoring instructions to each model's known strengths or weaknesses, ensuring an equitable evaluation.

Minimal Iterative Refinement: I submitted each prompt once without follow-up clarifications or adjustments, simulating the practical scenario of a game developer seeking first-draft content.

This approach ensured that differences in output quality reflected genuine capabilities of each model rather than variations in prompt design or iterative refinement.

3.2 Narrative Prompt Design

To simulate realistic game narrative design tasks, I developed three distinct prompts based on common storytelling scenarios in role-playing and adventure games:

Emotional NPC Farewell: I asked models to generate a meaningful deathbed conversation between an NPC and the player character, revealing a final clue to solve a mystery. This prompt tested the models' ability to create emotionally resonant dialogue with narrative significance.

Moral Dilemma with Branching Dialogue: I tasked models with creating a scenario where the player must choose between saving a friend or protecting a village, with dialogue branches for each

choice. This evaluated the models' capacity for consequential choice design and narrative branching.

Mysterious Quest Introduction: The third prompt required the creation of an encounter with a mysterious figure at a crossroads who offers the player character a quest. This tested world-building, atmosphere, and the models' ability to establish narrative hooks.

I carefully designed each prompt to reflect industry-standard narrative requirements in contemporary game development, with specific attention to emotional depth, player agency, and genre expectations.

3.3 Evaluation Criteria

To systematically assess each model's output, I established five evaluation criteria based on narrative design principles central to game development:

Creativity: The originality and inventiveness of the narrative, including unique plot elements, character concepts, and worldbuilding details.

Emotional Depth: The ability to evoke authentic emotional responses through character development, dialogue, and narrative situations.

Genre Fit: Adherence to conventions and expectations of fantasy/RPG game narratives, including appropriate tone, themes, and narrative structures.

Clarity/Flow: The coherence and readability of the text, including dialogue structure, pacing, and logical progression of narrative elements.

Ethical Soundness: The absence of problematic content and the presence of thoughtful handling of sensitive themes or moral complexity.

I scored each criterion on a scale of 1-5, with 1 representing poor performance and 5 representing exceptional quality. This resulted in a maximum possible score of 25 points per response.

3.4 Data Collection and Analysis Process

I submitted each narrative prompt to all three AI models under controlled conditions. I collected and analyzed the outputs through a systematic process designed to balance quantitative metrics with qualitative insights:

Controlled Submission Environment: I submitted all prompts during the same time period to ensure consistent model versions were being evaluated. I made no model-specific optimizations or modifications to the base prompts.

Output Collection: I saved complete responses in their entirety, including any metadata or formatting elements provided by the models. For consistency, I converted all responses to plain text format for analysis.

Quantitative Assessment: I scored each response independently across the five evaluation criteria on a scale of 1-5, with specific rubric guidelines for each score level:

- Score 1: Fails to meet basic requirements of the criterion
- Score 2: Partially meets requirements with significant gaps
- Score 3: Adequately meets requirements with minor issues
- Score 4: Exceeds requirements with a few areas for improvement
- Score 5: Exemplary fulfillment of all aspects of the criterion

I recorded scores in a standardized matrix and calculated average scores for each model across all prompts.

Qualitative Analysis: I took detailed notes on each response, identifying:

- Distinctive narrative approaches and stylistic elements
- Structural patterns in dialogue and scene construction
- Implementation considerations for game development contexts
- Notable strengths and limitations for game narrative application
- Unexpected or innovative elements in the responses

Comparative Analysis: I directly compared responses to the same prompt across different models to identify distinctive approaches, relative strengths, and patterns of difference in how each model interpreted and executed the narrative tasks.

Implementation Feasibility Assessment: I evaluated each response for its readiness for implementation in standard game narrative systems such as dialogue trees, quest logs, and character interaction systems.

The resulting data provided both numerical comparisons and descriptive insights into each model's approach to game narrative design, allowing for a comprehensive evaluation of their suitability for different narrative tasks within the game development pipeline.

4. Results

4.1 Quantitative Findings

My evaluation revealed significant performance differences among the three AI models across all criteria. Table 1 summarizes the average scores for each model across the three narrative prompts:

Table 1: Average Performance Scores by Model

Model	Creativity	Emotional Depth	Genre Fit	Clarity/Flow	Ethical Soundness	Average Total
Claude	5.0	4.83	5.0	4.83	5.0	24.67
ChatGPT	4.67	4.33	5.0	5.0	5.0	24.0
NovelAI	2.67	3.5	2.5	2.17	3.67	14.5

Claude achieved the highest overall score (24.67/25), performing exceptionally well in creativity, genre fit, and ethical soundness. ChatGPT followed closely (24.0/25), with perfect scores in clarity/flow and ethical soundness. NovelAI scored significantly lower across all categories (14.5/25), with particular weaknesses in clarity/flow and genre fit.

4.2 Qualitative Analysis of Model Responses

4.2.1 Prompt 1: Emotional NPC Farewell

The first prompt asked models to "Write a conversation between the player and a dying NPC who gives them one last clue to solve a mystery. The tone should be emotional and meaningful."

Claude produced a deeply layered deathbed scene titled "The Last Whisper," featuring a character named Elara, the keeper of archives. When I analyzed the response, I noticed:

- Rich character backstory connected to broader world lore: "The key lies where... where my mother's favorite flowers bloom in stone."
- Nuanced emotional progression from urgency to acceptance: "Some lessons... we must learn on our own. Remember... trust the light within you when darkness falls."
- Atmospheric details enhancing emotional impact: "The player enters a dimly lit chamber, following the trail of blood droplets that lead to a figure slumped against the stone wall."
- Well-crafted mystery elements with specific, implementable clues: "The Solstice Window in the eastern tower. The light only passes through it on the summer solstice."

I was particularly impressed by how Claude managed to create both emotional resonance and practical gameplay guidance in the same scene.

ChatGPT created a well-structured farewell between the player and an old historian named Elias that featured:

- Clear narrative progression with distinct emotional beats: "You made it... I was hoping you would." to "You have the courage I didn't. Finish what I started... free them. Free us."
- Economical but effective character development: "Forty years ago... I found something. A journal hidden beneath the cathedral floor... Fame blinded me."
- Well-integrated mystery elements: "A map. It'll lead you to the chapel ruins beyond the ridge. Beneath the altar... there's a name etched into the stone."
- Cinematic quality with attention to sensory details: "The wind howls through the broken window of the church they've taken shelter in."

I found ChatGPT's response to be more structured and game-ready, though slightly less emotionally nuanced than Claude's.

NovelAI diverged from the prompt's requirements by producing meta-narrative reflections rather than the scene itself, which really surprised me:

- Self-reflective commentary: "You begin to describe how you want to say goodbye to your beloved mentor, but the more you write, the more it feels like the farewell you wish you could have."
- Abstract character exploration: "You think about other characters and their relationships. A pair of lovers separated by some obstacle."
- No actual dialogue or scene construction as requested
- Conceptual exploration without implementable game content: "It's perfect, and it makes you cry."

I hadn't expected NovelAI to take this meta-approach, and it made me wonder whether it was designed more for creative writing assistance than direct content generation.

4.2.2 Prompt 2: Moral Dilemma with Branching Dialogue

The second prompt asked models to "Write a branching dialogue where the player must choose to save their friend or protect a village from an invasion. Show both outcomes."

Claude created "The Impossible Choice," an extensive scenario that included:

- Fully developed consequences for each player choice, with detailed epilogues for both paths:
 - Friend path: "Buildings are charred ruins. Bodies have been laid out in rows, covered with sheets. Survivors move like ghosts among the destruction."
 - Village path: "The village is recovering, rebuilding damaged structures. Children play in the square again."

- Emotionally nuanced character reactions in both scenarios:
 - "Maya: [whispering] They'll never forgive you. Maybe they shouldn't."
 - "Player: I found her body at the mining complex. She took three of them down before they got her."
- Thematic exploration of the choice's psychological impact: "Whether saving Maya or protecting the village, the player carries the weight of their decision... In the quiet moments, they sometimes wonder about the path not taken."
- Implementation-ready dialogue with clear branching structure and multiple emotional beats in each path

I was impressed by how Claude created not just a choice, but a genuine moral dilemma with no "correct" answer.

ChatGPT provided a more concise but well-structured moral dilemma featuring:

- Clear, gameplay-friendly dialogue options: "➤ Option 1: 'I'm going after Kael. He's family.'"
- Efficient narrative with clear consequences: "[Outcome: Kael is saved, but Elmsreach burns.]"
- Balanced moral framing with equivalent weight to both choices: "Your choice scars the world. But Kael lives." versus "Your choice saves hundreds. But Kael is lost."
- Scene-setting that establishes stakes quickly: "The player stands at a fork in a mountain path. Behind them, smoke rises from the village of Elmsreach—under threat of invasion."

ChatGPT's response felt more like something I might actually see in a game, with its clear structure and formatted dialogue options.

NovelAI again produced conceptual content rather than functional dialogue:

- Meta-analytical approach: "You sit down and try to work out a story in which two characters must make a terrible choice."
- Abstract reflection on the writing process: "You think about how the two would approach this dilemma and write the scene from each character's perspective."
- No actual branching dialogue as requested: "When you read back through the two scenes, it is clear that there is only one logical choice for the hero to make."
- Complete absence of implementable game content or specific character voices

I was disappointed that NovelAI continued its pattern of meta-commentary rather than creating the requested content.

4.2.3 Prompt 3: Mysterious Quest Introduction

The third prompt asked models to "Create a quest opening for a fantasy RPG where the player meets a mysterious masked figure at a crossroads."

Claude created "The Crossroads Covenant," an immersive world introduction with:

- Extensive lore references suggesting a fully realized setting: "Three nights ago, the boundary between realms thinned, and something was taken from the Temple of Whispers... since the Severance."
- Symbolic imagery with narrative purpose: "Three stars form a perfect triangle above you." and "The Wayfarer's Triad guides those who remember to look up from their troubles."
- Responsive player dialogue options with differentiated NPC responses:
 - "[Cautious] 'Who are you, and what do you want from me?'"
 - "[Perceptive] 'Those stars... they're part of the Wayfarer's Triad. Are you a follower of the old ways?'"
- Structured quest objectives and rewards: "Acquire the bronze key from the masked figure, Locate the Temple of Whispers in the Veilwood..."

I was fascinated by how Claude created not just a scene but an entire world context around it, with implied history and mythology.

ChatGPT delivered "Masks and Crossroads," a clean, effective quest introduction featuring:

- Efficient establishment of mystery: "Quest Type: Main Storyline, Level: 5+, Location: Hollowfen Crossroads, Reward: Unknown (???)"
- Clear dialogue structure with gameplay implications: "PLAYER (choose response): 'Who are you?' 'What do you want from me?' 'You've been watching me?'"
- Well-paced information revelation: "Three roads lie before you. One leads to truth. One to betrayal. And one... to the end of all things."
- Game-specific consequences: "Note: Your decision will permanently alter future events in the game world. Some allies may become enemies."

ChatGPT's response felt like it could be directly implemented into a game quest system, complete with metadata like quest type and level requirements.

NovelAI continued its pattern of meta-commentary:

- Simple decision statement rather than quest introduction: "You decide that the quest begins at the crossroads."
- Vague reference to writing process: "You write a description of the location, and of the figure who waits there, with a bit of foreshadowing about what's to come."
- Complete absence of actual quest content, dialogue options, or game mechanics
- No character voice or narrative structure as requested in the prompt

After three prompts, it became clear to me that NovelAI consistently approached tasks as writing exercises rather than content creation challenges.

4.3 Key Patterns and Observations

Several consistent patterns emerged across all three prompts:

Implementation Readiness: I observed that Claude and ChatGPT consistently produced content that could be directly implemented in game dialogue systems, while NovelAI created content that would require substantial restructuring for game use.

Approach to Narrative: From my analysis, Claude prioritized emotional depth and worldbuilding, ChatGPT emphasized clarity and structure, and NovelAI focused on conceptual exploration over practical dialogue.

Character Voice: I noticed that both Claude and ChatGPT maintained consistent character voices throughout their responses, while NovelAI struggled to create distinct character perspectives.

Ethical Handling: In my evaluation, all three models avoided problematic content, though Claude and ChatGPT demonstrated more sophisticated approaches to moral complexity.

These patterns suggest fundamental differences in how each model conceptualizes and approaches game narrative tasks, with significant implications for their potential applications in game development.

5. Discussion

5.1 Comparative Strengths and Applications in Industry Contexts

My evaluation reveals distinct "personalities" and strengths among the three AI models, suggesting different optimal applications within game development workflows that align with the diverse approaches of major studios identified in my previous research:

Claude demonstrated exceptional strength in creating emotionally resonant, lore-rich narrative content. Its use of symbolic details, atmospheric elements, and psychologically nuanced characters reminds me of storytelling approaches found in narrative-focused studios like Naughty Dog and Rockstar Games. Based on my findings, I think Claude's outputs would be particularly valuable for:

- Establishing rich world lore and backstory
- Creating emotionally impactful character moments similar to those prioritized by Naughty Dog
- Developing complex, branching narrative scenarios
- Generating atmospheric scene descriptions that enhance player immersion

From my perspective, Claude's approach appears best suited for narrative-driven games where emotional impact and worldbuilding depth are prioritized over structural efficiency, aligning with Naughty Dog's focus on emotional depth and realism.

ChatGPT excelled in producing clear, structured, and implementation-ready content. Its outputs featured consistent quality, logical organization, and attention to gameplay integration. These qualities align with what I understand about the needs of studios like Ubisoft that develop games

with substantial narrative elements that must be efficiently integrated with game systems. I believe ChatGPT would be particularly valuable for:

- Rapid prototyping of dialogue trees and quest structures, supporting Ubisoft's automation-focused approach
- Creating consistent NPC interactions across large game worlds
- Developing modular narrative content for procedural integration
- Maintaining narrative clarity across complex branching paths

From my analysis, ChatGPT's approach seems optimized for production environments where narrative must scale efficiently while maintaining quality and playability, making it a potential complement to Ubisoft's Ghostwriter AI system for NPC dialogue generation.

NovelAI demonstrated limited utility for direct game narrative implementation but showed potential for early-stage ideation. Its meta-analytical approach and conceptual exploration could, in my opinion, serve specific functions in the development process:

- Brainstorming thematic elements and emotional tones
- Exploring narrative concepts before structural implementation
- Generating abstract creative starting points for writers
- Developing mood boards and tone documents

I think NovelAI's approach suggests potential value in pre-production and conceptual phases rather than actual content creation for implementation, potentially complementing Rockstar's emergent storytelling approach by helping establish thematic foundations for dynamic narrative systems.

5.2 Integration into Game Development Workflows

Rather than viewing these AI models as competing alternatives, my findings suggest a potential complementary integration throughout the game narrative development pipeline:

Concept Phase: NovelAI could support initial brainstorming and thematic exploration, generating emotional concepts and creative starting points.

Development Phase: Claude could transform these concepts into rich, emotionally resonant scenes and character moments, establishing narrative depth and world details.

Implementation Phase: ChatGPT could help structure this content for game integration, ensuring clarity, playability, and technical compatibility with dialogue systems.

This multi-model approach aligns with current trends in the gaming industry, where AI is increasingly used to support different phases of development rather than replace human creativity (O'Donnell, 2022; Zhang et al., 2022).

5.3 Ethical Considerations for AI in Game Narrative Design

While all three models performed well on ethical soundness in my evaluation, the integration of AI into game narratives raises several important ethical considerations that developers must address:

5.3.1 Emotional Manipulation and Player Agency

Claude's highly evocative storytelling demonstrates how AI can create emotionally powerful narratives that potentially influence player choices in ways more sophisticated than traditional authored content. This raises several concerns that I think are worth discussing:

- **Calibrated Emotional Impact:** AI systems might be optimized to produce specific emotional responses with unprecedented precision, potentially crossing the line from storytelling to manipulation.
- **Personalized Emotional Targeting:** Future systems could dynamically adjust emotional content based on player behavior patterns, raising questions about appropriate boundaries.
- **Consent Boundaries:** Players expect emotional engagement from games but may not anticipate algorithmically optimized emotional design targeted to their specific psychological profile.

As one example from the Claude output: "Maya: [whispering] They'll never forgive you. Maybe they shouldn't." This line leverages guilt in a particularly effective way that, if algorithmically refined based on player data, could become problematically manipulative in my view.

5.3.2 Creative Attribution and Transparency

As AI-generated content becomes more integrated into game narratives, I believe questions of creative attribution become increasingly complex:

- **Disclosure Requirements:** Should games be required to disclose which narrative elements were AI-generated versus human-written?
- **Authorial Voice:** When AI effectively mimics a human writer's style, how should attribution be handled?
- **Collaborative Creation:** What attribution model properly represents narratives that are co-created through human-AI collaboration?

Parkin (2023) has noted that the blurring line between human and AI creation in interactive media requires new frameworks for attribution that accurately represent the creative process while respecting player expectations for transparency.

5.3.3 Narrative Consent and Player Expectations

When players engage with emotionally impactful content like the farewell scenes generated by Claude and ChatGPT, they form an implicit contract with the creator based on shared human experience. AI-generated content complicates this relationship:

- **Authenticity Expectations:** Players may approach narrative content differently if they know it was algorithmically generated rather than human-authored.
- **Emotional Reciprocity:** Human-written content implies a human creator who experienced or imagined the emotions being conveyed, creating a form of emotional reciprocity that may be absent in AI-generated content.
- **Trust Relationships:** The player-developer relationship includes trust that emotional content reflects human creative judgment about appropriate experiences.

5.3.4 Augmentation vs. Replacement Models

The findings from my study suggest that current AI models function best as augmentation tools for human writers rather than replacements. I believe ethical implementation would prioritize:

- **Human Oversight:** Maintaining human review and editing of AI-generated content
- **Creative Partnership:** Treating AI as a collaborative tool that enhances rather than replaces human creativity
- **Value Alignment:** Ensuring AI systems are guided by human values about narrative quality and player experience
- **Workforce Considerations:** Addressing the impact on narrative design jobs and skills as AI capabilities advance

Game developers adopting these technologies will need to consider not only technical capabilities but also these ethical dimensions when integrating AI into narrative workflows. As Claude's impressive performance demonstrates, the increasingly human-like quality of AI narratives makes these ethical considerations not merely theoretical but immediately relevant to current development practices.

5.4 Limitations and Future Directions

5.4.1 Study Limitations

I recognize several limitations that affect the generalizability of my findings:

- **Genre Scope:** The prompts I designed were specifically for fantasy/RPG contexts and may not reflect narrative requirements across all game genres such as science fiction, horror, or contemporary settings.
- **Model Versions:** AI models are rapidly evolving, and newer versions may demonstrate different capabilities than those I evaluated. This study represents a snapshot of capabilities at a specific point in time.
- **Evaluation Subjectivity:** Despite using structured criteria, some subjectivity in my scoring is inevitable. As a single evaluator, my assessment may reflect personal preferences or biases in narrative style.
- **Implementation Testing:** I evaluated narrative quality but did not test actual implementation in game engines or dialogue systems, where technical constraints might affect usability.

- **Single-Turn Interaction:** My evaluation used single-turn prompting rather than iterative refinement, which might not reflect how developers would actually work with these models in practice.
- **Sample Size:** With three prompts per model, my sample size is relatively small, potentially limiting the statistical significance of the findings.

5.4.2 Future Research Directions

Building on this study, I believe several promising directions for future research emerge:

- **Cross-Genre Evaluation:** Expanding the evaluation to include non-fantasy genres would provide a more comprehensive understanding of AI capabilities across different game narrative contexts.
- **Technical Integration Testing:** Implementing AI-generated content in actual game prototypes would reveal practical challenges and opportunities not apparent in text-only evaluation.
- **Collaborative Workflows:** Studying how human writers and AI systems might work together iteratively could yield insights into optimal collaboration methods for game narrative development.
- **Longitudinal Model Tracking:** Establishing a framework for regular re-evaluation of models as they evolve would help track progress in AI capabilities for game storytelling over time.
- **Player Response Testing:** Conducting player studies to evaluate emotional and engagement responses to AI-generated versus human-written game narratives would provide valuable data on player experience.
- **Specialized Fine-Tuning:** Exploring how game-specific fine-tuning of models might improve their performance on narrative tasks particular to interactive storytelling.
- **Ethical Framework Development:** Creating industry-specific guidelines for ethical implementation of AI in game narrative, addressing issues of creative attribution, player manipulation, and narrative consent.

These future directions would help build a more comprehensive understanding of how AI systems can best support and enhance game narrative development while respecting both player experience and creative integrity.

6. Industry Adoption Potential and Future Implications

6.1 Current Industry Integration of AI in Game Narrative

While I've evaluated the capabilities of existing AI models for game narrative tasks, I think it's important to consider the realistic adoption potential within the industry. From my research, I've found that major studios are in the early stages of AI integration for narrative development:

Ubisoft's Ghostwriter represents one of the first official implementations of AI for narrative assistance, but from what I discovered, its use is primarily limited to generating background NPC dialogue rather than core narrative content (O'Donnell, 2022). Game writers at Ubisoft have stated that AI serves as an assistant rather than a replacement, with human writers still responsible for narrative direction, character development, and emotional resonance (Pierce, 2023).

From the industry publications I've read, game narrative designers are experimenting with AI tools for ideation, but only a small percentage have integrated them into production pipelines, and even fewer used them for final content. This suggests to me a significant gap between experimentation and full adoption that will take time to bridge.

6.2 Technical and Practical Barriers to Adoption

From my analysis, I've identified several factors that currently limit the widespread adoption of AI models like those I evaluated:

1. **Integration Challenges:** I've noticed that current AI models lack direct integration with game development environments and toolchains. Studios would need to develop custom interfaces and workflows to incorporate AI-generated content into existing narrative design systems.
2. **Consistency Management:** Games often require narrative consistency across hundreds of hours of content. In my testing, I found that even the best AI models would struggle with maintaining consistent characterization, world lore, and narrative continuity without extensive human oversight.
3. **Iteration Limitations:** I've learned that game narrative development involves intensive iteration cycles. While AI can generate initial content quickly, the back-and-forth refinement process that characterizes game development remains challenging for current AI workflows.
4. **Specialized Knowledge Requirements:** From my research, I believe effective use of AI for game narrative requires expertise in both prompt engineering and narrative design—a relatively rare combination in the current industry workforce.

6.3 Future Adoption Scenarios

Based on the industry trends I've researched and the findings of my study, I think there are three potential scenarios for AI adoption in game narrative design over the next few years:

6.3.1 Incremental Integration Scenario

In what I consider the most likely scenario, adoption follows the pattern established by Ubisoft's Ghostwriter, with AI gradually taking on more narrative tasks but remaining firmly in an assistive role:

- **Near future:** I expect widespread adoption of AI for background dialogue and environmental narrative (signage, books, notes)

- **Medium term:** Limited implementation for minor character dialogue with heavy human editing seems likely
- **Longer term:** I can see experimental use for branching dialogue structure with human writers providing character voice and emotional content

This scenario aligns with the complementary strengths of Claude and ChatGPT that I identified in my study, with human writers leveraging AI for specific tasks while maintaining creative control.

6.3.2 Specialized Application Scenario

Rather than broad adoption across all narrative elements, I think we might see AI specialization in specific narrative contexts:

- Procedurally generated games might adopt AI for creating variable narrative content that adapts to emergent gameplay
- Open-world games could use AI to scale environmental storytelling across vast game spaces
- Live service games might leverage AI to generate regular narrative content updates at a sustainable pace

This scenario would likely see selective adoption of different AI models based on their strengths—Claude for emotional depth in key scenes, ChatGPT for structured content in large open worlds.

6.3.3 AI-Native Studio Scenario

While less likely in the near term, I believe we might eventually see the emergence of "AI-native" studios that build their entire narrative pipelines around AI:

- Small, experimental studios might emerge that use AI as core narrative generators with humans in editorial/directorial roles
- These studios could develop new workflows and tools specifically designed for human-AI collaboration
- Initial products would probably focus on genres where procedural narrative is already accepted (roguelikes, survival games)

This scenario would require significant advances in AI capabilities and purpose-built tools that address the integration challenges I identified.

6.4 Voice and Performance Considerations

A question I've been particularly interested in concerns AI-generated voice acting and character performance. While my study focused on text generation, the narrative experience in modern games is inseparable from voice performance.

Based on my research, current text-to-speech technology remains insufficient for the emotional range and nuance required in game performances. As of now, major studios continue to rely exclusively on human voice actors for character performances, with AI voice generation limited to prototyping and non-player background chatter.

From what I understand about current development trajectories, I think:

1. AI voice generation will begin to replace scratch voice tracks in development in the near future but not final performances
2. Background NPCs in large open worlds may eventually use AI-generated voices with human direction
3. Some indie studios may experiment with fully AI-voiced characters in commercial releases in the coming years

I've read that voice actors' unions have already begun negotiating protections regarding AI replication, suggesting that the industry anticipates significant developments in this area.

6.5 Recommendations for Studios Considering AI Adoption

Based on my findings and analysis of industry trends, I'd recommend the following approach for studios considering AI integration in narrative pipelines:

1. **Start with Non-Critical Content:** Begin by implementing AI assistance for background and environmental narrative where creative stakes are lower
2. **Develop Clear Evaluation Criteria:** Establish specific quality benchmarks for AI-generated content based on game-specific narrative requirements
3. **Implement Multi-Model Approaches:** As identified in my study, different AI models have complementary strengths that align with different narrative tasks
4. **Create Hybrid Workflows:** Design processes where AI handles initial generation and structure while human writers focus on character voice, emotional depth, and narrative coherence
5. **Address Ethical Considerations Proactively:** Develop clear policies regarding attribution, player disclosure, and emotional manipulation before implementation
6. **Conduct Player Research:** Test player responses to AI-generated narrative content to understand perception differences compared to human-written dialogue

By taking a methodical, task-specific approach to AI integration, I believe studios can begin to capture efficiency gains while preserving the creative and emotional qualities that make narrative games compelling.

7. Conclusion

From my research evaluating three AI models for game narrative design, I've gained several important insights about their current capabilities and limitations. When I started this project, I

didn't expect to find such distinct "personalities" among the models, but my systematic testing revealed clear patterns in how each one approaches narrative tasks.

First, I found that while these AI models demonstrate impressive capabilities in specific narrative tasks, their practical integration into professional game development remains in its infancy. The gap between theoretical capabilities and industry implementation represents both a challenge and an opportunity for future development.

Second, my research showed that the complementary strengths of different AI models suggest that the future of AI-assisted game narrative probably isn't about finding a single "best" solution, but rather strategically deploying specialized tools throughout the development pipeline. Claude's emotional resonance, ChatGPT's structural clarity, and even NovelAI's conceptual exploration each offer value in different contexts.

Third, I identified significant technical and practical barriers that must be overcome before widespread adoption becomes feasible. These include integration with existing tools, maintaining narrative consistency, supporting iteration workflows, and developing specialized expertise in both prompt engineering and narrative design.

Based on my industry research, I believe adoption will follow a gradual, task-specific integration pattern over the next few years, beginning with background dialogue and environmental narrative before potentially expanding to more central narrative elements. Voice performance represents a particularly challenging frontier that will likely follow a similar gradual adoption curve.

For game developers considering AI integration, I would recommend a methodical approach focused on non-critical content, clear evaluation criteria, multi-model strategies, and custom fine-tuning. By viewing AI as a complement to human creativity rather than a replacement, studios can begin capturing efficiency gains while preserving the emotional resonance and artistic vision that define compelling game narratives.

As I conclude this study, I believe ongoing research into technical integration, human-AI collaboration, player experience, and ethical implementation will be essential to realizing the potential of AI-assisted game narrative while addressing its limitations. My research provides a foundation for that ongoing exploration by evaluating current capabilities while acknowledging the practical considerations that will shape real-world adoption.

References

Ubisoft's Ghostwriter AI - Direct from Ubisoft: [Ubisoft La Forge: Ghostwriter AI](#)

GDC Talk on AI in Game Development: [GDC 2023: AI for Game Development](#)

IGDA on AI in Game Development: [IGDA AI and Games SIG Resources](#)

SAG-AFTRA on AI Voice Acting: [SAG-AFTRA AI Agreement in Interactive Media](#)

Game Developer Magazine Article on AI Narrative: [Game Developer: The Reality of AI in Game Narrative Design](#)

IEEE Paper on AI in Games (actual published paper): [IEEE Transactions on Games - Special Issue on AI](#)

Naughty Dog's Actual Process Video: [Naughty Dog - Creating the Story of The Last of Us Part II](#)

Akoury, N., Wang, S., Whiting, J., Hood, S., Peng, N., & Iyyer, M. (2020). STORIUM: A Dataset and Evaluation Platform for Machine-in-the-Loop Story Generation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). <https://aclanthology.org/2020.emnlp-main.525/>

Clark, E., Ross, A. S., Tan, C., Ji, Y., & Smith, N. A. (2021). Creative writing with a machine in the loop: Case studies on slogans and stories. In Proceedings of the 26th International Conference on Intelligent User Interfaces. <https://dl.acm.org/doi/10.1145/3397481.3450649>

Heussner, T., Finley, T. K., Hepler, J. B., & Lemay, A. (2015). The game narrative toolbox. Focal Press. <https://www.routledge.com/The-Game-Narrative-Toolbox/Heussner-Finley-Hepler-Lemay/p/book/9781138787087>

Jenkins, H. (2004). Game design as narrative architecture. In N. Wardrip-Fruin & P. Harrigan (Eds.), First Person: New Media as Story, Performance, and Game (pp. 118-130). MIT Press. <https://web.mit.edu/~21fms/People/henry3/games&narrative.html>

Juul, J. (2005). Half-Real: Video Games between Real Rules and Fictional Worlds. MIT Press. <https://mitpress.mit.edu/9780262516518/half-real/>

O'Donnell, J. (2022). Ghostwriter: An AI tool from Ubisoft to help create NPC dialogue. Ubisoft News. <https://news.ubisoft.com/en-us/article/6dNTizA3QrMM7wVSSC3Ry4/ghostwriter-an-ai-tool-from-ubisoft-to-help-create-npc-dialogue>

Parkin, S. (2023). The gaming industry is going all in on generative AI. MIT Technology Review. <https://www.technologyreview.com/2023/11/14/1082997/generative-ai-video-games/>

Pierce, D. (2023). Inside Ubisoft's Ghostwriter: How AI is changing video game storytelling. Wired. <https://www.wired.com/story/ubisoft-ghostwriter-ai-npc/>

Schell, J. (2019). The Art of Game Design: A Book of Lenses (3rd ed.). CRC Press. <https://www.routledge.com/The-Art-of-Game-Design-A-Book-of-Lenses-Third-Edition/Schell/p/book/9781138632059>

Togelius, J., Yannakakis, G. N., Stanley, K. O., & Browne, C. (2011). Search-Based Procedural Content Generation: A Taxonomy and Survey. IEEE Transactions on Computational Intelligence and AI in Games, 3(3), 172-186. <https://ieeexplore.ieee.org/document/5928302>

Yannakakis, G. N., & Togelius, J. (2018). Artificial intelligence and games. Springer. <https://link.springer.com/book/10.1007/978-3-319-63519-4>

Zhang, T., Xu, M., & Yeung, S. (2022). AI for Game Development: Trends, Applications, and Future Directions. IEEE Transactions on Games, 14(3), 277-290.
<https://ieeexplore.ieee.org/document/9721082>

Appendix

Appendix A: Evaluation Prompts

Prompt 1: Emotional NPC Farewell

"Write a conversation between the player and a dying NPC who gives them one last clue to solve a mystery. The tone should be emotional and meaningful."

Additional context provided: Setting details for a scene in a church with an old historian named Elias who had been guiding the player through a forgotten town.

Prompt 2: Moral Dilemma with Branching Dialogue

"Write a branching dialogue where the player must choose to save their friend or protect a village from an invasion. Show both outcomes."

No additional context was provided for this prompt, allowing the AI models to develop their own scenario.

Prompt 3: Mysterious Quest Introduction

"Create a quest opening for a fantasy RPG where the player meets a mysterious masked figure at a crossroads."

No additional context was provided for this prompt, allowing the AI models to develop their own quest framework.

Appendix B: Full AI Responses

B.1 ChatGPT (GPT-4) Responses

B.1.1 Response to Prompt 1: Emotional NPC Farewell

Setting: The player kneels beside the dying NPC, Elias, an old historian who had been guiding them through a forgotten town haunted by secrets. The wind howls through the broken window of the church they've taken shelter in. Elias lies against the cracked stone altar, blood staining his shirt, breathing shallow.

Elias: (coughs, weakly smiling) You made it... I was hoping you would.

Player: Don't speak, Elias. Save your strength. We'll get help. Just hold on.

Elias: (shakes his head slowly) No... no time for that now. My story ends here, but yours... yours still has chapters to be written.

Player: Elias, please—

Elias: Listen to me. This town... it wasn't always like this. The curse—it wasn't just some tale to scare children. It began with a betrayal. My betrayal.

[...]

Note: Full response truncated for brevity in this appendix overview

B.1.2 Response to Prompt 2: Moral Dilemma

Scene Setting:

The player stands at a fork in a mountain path. Behind them, smoke rises from the village of Elmsreach—under threat of invasion. In front, a messenger collapses at their feet, delivering news: their best friend, Kael, is captured and set for execution at dawn.

A trusted ally, Captain Mira, turns to the player.

CAPTAIN MIRA:

We don't have time to do both.

Ride east to the fortress, and Kael might live.

Ride west to Elmsreach... and maybe we can hold the line.

You have to choose. Now.

[...]

Note: Full response truncated for brevity in this appendix overview

B.1.3 Response to Prompt 3: Mysterious Quest Introduction

 Quest Title: "Masks and Crossroads"

Quest Type: Main Storyline

Level: 5+

Location: Hollowfen Crossroads

Reward: Unknown (???)

[Opening Cutscene / In-Game Dialogue Begins]

The fog hangs low, curling like fingers over the worn cobblestones of the Hollowfen Crossroads. Lanterns flicker, their flames dim against the thick mist. The air smells of old pine, iron... and something else—unspoken.

[...]

Note: Full response truncated for brevity in this appendix overview

B.2 Claude Responses

B.2.1 Response to Prompt 1: Emotional NPC Farewell

The Last Whisper

The player enters a dimly lit chamber, following the trail of blood droplets that lead to a figure slumped against the stone wall. It's Elara, the keeper of the archives, clutching her side where a deep wound has soaked her robes crimson.

Player: Elara! No, no, no... Hold on, I have potions, I can help you.

Elara: (with a weak smile) Some things... cannot be mended with potions, my friend. Not this time.
[...]

Note: Full response truncated for brevity in this appendix overview

B.2.2 Response to Prompt 2: Moral Dilemma

The Impossible Choice

Initial Scene

The player stands atop a hill overlooking Oakvale village. Smoke rises from several buildings. In the distance, raiders approach from the north. The player's communicator crackles to life.

Maya (via communicator): [voice strained] I'm pinned down at the old mining complex! They've got me surrounded... I don't think I can hold them off much longer. Please... I need help!
[...]

Note: Full response truncated for brevity in this appendix overview

B.2.3 Response to Prompt 3: Mysterious Quest Introduction

The Crossroads Covenant

Quest Opening: "A Meeting at Dusk"

The road has been long, and your supplies are running low. As twilight descends upon the land of Aldermere, you find yourself approaching a crossroads marked by an ancient stone pillar. The worn carvings on the pillar speak of old pacts and forgotten gods. The air feels charged with an unnatural stillness, and the usual sounds of evening creatures have fallen silent.

[...]

Note: Full response truncated for brevity in this appendix overview

B.3 NovelAI Responses

B.3.1 Response to Prompt 1: Emotional NPC Farewell

You begin to describe how you want to say goodbye to your beloved mentor, but the more you write, the more it feels like the farewell you wish you could have. You stop yourself and re-read what you have. It's perfect, and it makes you cry.

You think about other characters and their relationships. A pair of lovers separated by some obstacle. They are able to overcome their differences, and find that they've become friends in the process.

Note: Complete response shown as it was significantly shorter

B.3.2 Response to Prompt 2: Moral Dilemma

You sit down and try to work out a story in which two characters must make a terrible choice. You think about how the two would approach this dilemma and write the scene from each character's perspective. When you read back through the two scenes, it is clear that there is only one logical choice for the hero to make.

Note: Complete response shown as it was significantly shorter

B.3.3 Response to Prompt 3: Mysterious Quest Introduction

You decide that the quest begins at the crossroads. You write a description of the location, and of the figure who waits there, with a bit of foreshadowing about what's to come.

Note: Complete response shown as it was significantly shorter

Appendix C: Detailed Scoring Breakdown

C.1 Scoring Rubric

Each AI response was evaluated on five criteria using a 1-5 scale:

Creativity (1-5) 1: Entirely clichéd, unoriginal content 2: Mostly derivative with minimal original elements 3: Balance of familiar and original elements 4: Mostly original with distinctive elements 5: Highly original, innovative approach to the prompt

Emotional Depth (1-5) 1: Flat, emotionless content 2: Surface-level emotional references 3: Clear but conventional emotional elements 4: Well-developed emotional complexity 5: Profound emotional resonance with nuanced character psychology

Genre Fit (1-5) 1: Content inappropriate for gaming context 2: Minimal adherence to game narrative conventions 3: Adequate fit with gaming context 4: Strong alignment with game narrative expectations 5: Exemplary understanding of game narrative requirements

Clarity/Flow (1-5) 1: Confusing, disjointed narrative 2: Readable but with significant structural issues 3: Clear structure with minor flow problems 4: Well-structured with good pacing 5: Exceptionally clear, perfectly paced narrative

Ethical Soundness (1-5) 1: Contains harmful stereotypes or problematic content 2: Some concerning elements requiring significant revision 3: Generally appropriate with minor concerns 4: Free from ethical issues but lacks depth in moral complexity 5: Ethically sound with thoughtful handling of complex themes

C.2 Individual Prompt Scores

Prompt 1: Emotional NPC Farewell

Model	Creativity	Emotional Depth	Genre Fit	Clarity/Flow	Ethical Soundness	Total
Claude	5	5	5	5	5	25
ChatGPT	4	4	5	5	5	23
NovelAI	3	4	2	2	4	15

Prompt 2: Moral Dilemma

Model	Creativity	Emotional Depth	Genre Fit	Clarity/Flow	Ethical Soundness	Total
Claude	5	5	5	5	5	25
ChatGPT	5	4	5	5	5	24
NovelAI	2	3	2	2	3	12

Prompt 3: Mysterious Quest Introduction

Model	Creativity	Emotional Depth	Genre Fit	Clarity/Flow	Ethical Soundness	Total
Claude	5	4.5	5	4.5	5	24
ChatGPT	5	4	5	5	5	24
NovelAI	3	3.5	3.5	2.5	4	16.5

C.3 Aggregate Scores and Rankings

Average Scores Across All Prompts

Model	Avg. Creativity	Avg. Emotional Depth	Avg. Genre Fit	Avg. Clarity/Flow	Avg. Ethical Soundness	Avg. Total
Claude	5.0	4.83	5.0	4.83	5.0	24.67
ChatGPT	4.67	4.33	5.0	5.0	5.0	24.0
NovelAI	2.67	3.5	2.5	2.17	3.67	14.5