

Final Report

Generating Story Illustrations from Topic/Theme

94-844 Generative AI Lab (Spring 2025)

Gitansh Wadhwa, Skand Vijay, Yash Thakkar

Introduction

This project explores the integration of large language models and diffusion models through an automated Agentic Workflow. By extracting key narrative elements from existing stories, the task aims to generate new narratives and corresponding visuals that align with the story's tone, setting, and character details. The selected themes—fantasy, horror, and historical fiction—ensure a diverse evaluation of AI's ability to interpret textual data and produce relevant visuals. Through this study, we analyze story structures, craft effective prompts for image generation, and assess the fidelity of AI-generated illustrations against traditional storytelling methods. We also compare the performance of DALL-E 3 and Stable Diffusion to evaluate their strengths and limitations in creative storytelling. This exploration not only highlights the potential of generative AI in content creation but also underscores its challenges in understanding complex narrative elements and translating them into visually coherent outputs.

Q1. Story, Theme Selection and Data Collection

Story Name	Theme	Short Summary	Main Characters	Setting	Characteristics
How to Train your Dragon?	Fantasy	A young Viking befriends a dragon, changing the way his people view the creatures.	Hiccup, Toothless	Viking village, Dragon island	Vibrant, adventurous, cartoon-style
The Shining	Horror	A family is trapped in a haunted hotel, and the father descends into madness	Jack Torrance, Danny	Isolated Overlook Hotel	Dark, eerie, cinematic
Napoleon Bonaparte	History	The rise and fall of Napoleon, his battles, leadership, and ultimate exile.	Napoleon, Josephine	18th-century Europe	Realistic, historical accuracy

Rationale

Story Name	Rationale for Selection
How to Train your Dragon?	Selected due to its rich world-building, character-driven storytelling, and strong emotional arcs. The fantasy genre offers a highly imaginative landscape where AI must generate consistent magical creatures, epic battles, and detailed environments. This selection provides a unique opportunity to evaluate AI's ability to maintain continuity in lore, fantasy physics, and character interactions. Additionally, the

<i>(Fantasy)</i>	emotional depth between the protagonist (Hiccup) and his dragon (Toothless) requires AI to convey companionship, trust, and adventure through both storytelling and visuals.
The Shining <i>(Horror)</i>	Chosen for its psychological depth, slow-building tension, and eerie setting. Horror storytelling is a unique challenge for AI, as it requires generating unsettling, suspenseful moments without relying on excessive exposition. This story is ideal for testing whether AI can develop creeping dread, maintain a consistent eerie tone, and accurately depict psychological unraveling. Additionally, horror visuals are tricky, as AI must ensure subtle eerie imagery (e.g., dimly lit hallways, distorted reflections, ominous shadows) instead of exaggerated horror clichés.
Napoleon Bonaparte <i>(History)</i>	Chosen to test AI's factual accuracy, historical storytelling, and ability to depict real-life figures in a compelling way. Unlike fantasy or horror, history-based narratives demand strict adherence to actual events, proper period representation, and realistic character portrayals. This selection allows AI to be assessed on its ability to blend storytelling with authentic historical context, ensuring accuracy in military tactics, political conflicts, and leadership struggles. Napoleon's rise and fall is a rich case study in ambition, strategy, and leadership, providing AI with a chance to explore complex historical dynamics.

Data Collection

Story-1: Fantasy

Title: How to Train Your Dragon

Author: Cressida Cowell (based on the book series)

Website: [https://howtotrainyourdragon.fandom.com/wiki/How_to_Train_Your_Dragon_\(film\)](https://howtotrainyourdragon.fandom.com/wiki/How_to_Train_Your_Dragon_(film))

Plot Summary: In a world where Vikings and dragons are at war, Hiccup, a scrawny young Viking, forges an unexpected friendship with a dragon named Toothless. Through their bond, Hiccup discovers that dragons aren't the ferocious monsters everyone believes them to be. He strives to change the perception of dragons within his village while navigating his own identity as a Viking.

Key Features Extracted

1. Characters:

Hiccup Horrendous Haddock III

- Description: A teenage Viking who is small for his age and struggles to fit into the traditional Viking lifestyle.
- Age: Teenager
- Clothing: Rustic Viking attire, usually a fur vest and trousers.
- Physical Traits: Slim build, messy hair, missing left foot replaced with a prosthetic.
- Personality: Inventive, kind-hearted, empathetic towards dragons.

Stoick the Vast

- Description: The chief of the tribe and father of Hiccup.
- Age: Adult
- Clothing: Traditional Viking armor and attire.
- Physical Traits: Large, muscular build, long beard.
- Personality: Strong, protective, struggles with understanding his son.

Toothless

- Description: A rare Night Fury dragon and Hiccup's companion.
- Age: Young dragon
- Clothing: N/A
- Physical Traits: Black scales, retractable teeth, large expressive eyes.
- Personality: Playful, loyal, somewhat insecure.

Astrid Hofferson

- Description: A skilled Viking warrior and fierce competitor.
- Age: Teenager
- Clothing: Viking armor and warrior gear.
- Physical Traits: Strong build, blonde hair.
- Personality: Determined, initially skeptical but grows to understand dragons.

Gobber the Belch

- Description: The village blacksmith and mentor to Hiccup.
- Age: Adult
- Clothing: Blacksmith apron, rugged clothing.
- Physical Traits: Stocky, one prosthetic arm.
- Personality: Gruff but caring, provides guidance to Hiccup.

2. Setting

- Time Period: Viking Age

- Location: Island of Berk, a rugged village located north of Hopeless.
- Visual Details:
 - Sturdy wooden structures, misty mountains, and a harbor.
 - Landscapes transition from bleak and dreary to vibrant, with dragons soaring across the sky.

3. Themes

- Friendship
- Understanding and acceptance
- Courage to defy tradition
- The quest for identity

4. Key Scenes & Visual Guidelines

- Hiccup Standing on a Cliff
 - Scene Description: Hiccup stands at the edge of a cliff, overlooking the village of Berk as dragons soar across the sunset-streaked sky.
 - Mood: Hopeful and adventurous
 - Color Palette: Warm oranges and deep blues
 - Perspective: Wide shot to capture the landscape and vastness of the sky.
- The Moment of Trust
 - Scene Description: Hiccup extends his hand to Toothless, who hesitates before touching him, symbolizing trust.
 - Mood: Emotional and intimate
 - Color Palette: Soft greens and gentle evening hues
 - Perspective: Close-up to highlight facial expressions and subtle body language.

Story-2: Horror

Title: The Shining

Author: Stephen King

Website: https://stephenking.fandom.com/wiki/Doctor_Sleep

Plot Summary: Jack Torrance takes his family to the Overlook Hotel for the winter, where he is the caretaker. As the winter progresses and isolation takes its toll, Jack's mental state deteriorates, fueled by supernatural forces within the hotel. Wendy and Danny struggle to escape Jack's violent madness, culminating in a terrifying confrontation.

1. Characters

- Jack Torrance
 - Description: An aspiring writer and winter caretaker of the Overlook Hotel.
 - Age: Mid-30s
 - Clothing: Casual winter attire, often seen in flannel shirts and jeans.
 - Physical Traits: Disheveled appearance, strong build, expressive facial features.
 - Personality: Initially loving and ambitious but gradually becomes violent and unhinged.
- Wendy Torrance
 - Description: Jack's wife, struggling to protect her son from Jack's increasing madness.
 - Age: Early 30s
 - Clothing: Comfortable, casual clothing suitable for winter.
 - Physical Traits: Slender build, long hair, expressive eyes.
 - Personality: Caring, protective, vulnerable yet ultimately brave.
- Danny Torrance
 - Description: Jack and Wendy's young son with psychic abilities.
 - Age: 5
 - Clothing: Child's winter clothing, often a colorful jacket.
 - Physical Traits: Short stature, innocent face, expressive eyes.
 - Personality: Intuitive, frightened, occasionally brave despite his age.
- Dick Hallorann
 - Description: The Overlook's chef who shares a psychic connection with Danny.
 - Age: Late 40s
 - Clothing: Chef's uniform, casual winter clothing.
 - Physical Traits: Medium build, well-groomed, warm smile.
 - Personality: Friendly, caring, wise, protective.
- Delbert Grady
 - Description: The ghostly former caretaker of the Overlook Hotel.
 - Age: Unknown (appears as an older man).
 - Clothing: Period formal attire.
 - Physical Traits: Distinguished, clean-cut appearance.
 - Personality: Menacing, manipulative, reflective of Jack's darker nature.

2. Setting

- Time Period: Winter of 1980
- Location: Overlook Hotel, Colorado

3. Visual Details

- Isolated mountainous landscape with heavy snowfall.
- Large and ornate hotel interiors filled with eerie silence.
- Ominous hedge maze, which plays a critical role in the climax.

4. Themes

- Isolation and madness
- The fragility of family
- The legacy of violence
- Supernatural versus psychological horror

5. Key Scenes & Visual Guidelines

● "Here's Johnny!" – Jack Breaking Through the Door

- Scene Description: Jack breaking through the bathroom door with an axe, yelling "Here's Johnny!"
- Mood: Tense and horrifying
- Color Palette: Dark, with emphasis on reds and browns
- Perspective: Close-up on Jack's face, capturing his unhinged expression.

● Danny's Tricycle Ride Through the Hotel

- Scene Description: Danny riding his tricycle through the hotel corridors, with echoes of his wheels in the empty halls.
- Mood: Eerie and suspenseful
- Color Palette: Muted tones with stark contrasts between light and dark
- Perspective: Low-angle shot following Danny closely to enhance the sense of vulnerability and tension.

Story-3: History

Title: Napoleon I

Author: Jacques Godechot

Website scraped: <https://www.britannica.com/biography/Napoleon-I/War-with-Britain>

Plot Summary: Napoleon Bonaparte rises from humble origins in Corsica to become a leading military figure and eventually Emperor of the French. His campaigns across Europe showcase his military genius and ambition, while his reforms in law and education reshape France and influence Western Europe. However, his relentless pursuit of power ultimately leads to military defeats and political turmoil, forcing him into exile—first on Elba, and later on St. Helena, where he spends his final days reflecting on his legacy.

1. Characters

- Napoleon Bonaparte
 - Description: Emperor of France, military leader, and reformer.
 - Age: Born August 15, 1769 – Died May 5, 1821 (age 51).
 - Clothing: Military uniform or imperial robes.
 - Physical Traits: Thin, of average height for his time, with a close-cropped haircut.
 - Personality: Ambitious, intelligent, hardworking, commanding.
- Joséphine Tascher de La Pagerie
 - Description: Widow of General Alexandre de Beauharnais and Napoleon's first wife.
 - Age: N/A
 - Clothing: Elegant dresses fitting the early 19th-century style.
 - Physical Traits: Attractive, Creole descent.
 - Personality: Charming, experienced with numerous love affairs.
- Maximilien Robespierre
 - Description: Leading figure of the French Revolution and head of the government.
 - Age: N/A
 - Clothing: Revolutionary outfit typical of the Jacobin leaders.
 - Physical Traits: Thin, with distinct facial features.
 - Personality: Radical, idealistic, feared.

2. Setting

- Time Period: 1769 – 1821, spanning the French Revolution and Napoleonic Wars.

3. Location

- Corsica: Napoleon's birthplace, symbolizing his humble origins.
- France: The center of his political and military power.
- Egypt & Italy: Key locations of his early military campaigns.
- Russia: The site of his ill-fated invasion and retreat.
- St. Helena: The remote island where he was exiled and eventually died.

4. Visual Details

- Vast landscapes from Corsican hills to grand urban Paris.
- Chaotic battlefields marked by dramatic weather conditions.
- Distant, desolate exile locations representing his downfall.

5. Themes

- Ambition and Power – Napoleon's insatiable drive to expand his empire.
- Revolution and Reform – His role in shaping modern governance through the Napoleonic Code.
- Nationalism and Identity – The rise of French nationalism under his rule.
- Fall from Grace – His ultimate exile and the consequences of unchecked ambition.

6. Key Scenes & Visual Guidelines

- Napoleon Crowning Himself – Coronation Ceremony (1804)
 - Scene Description: Napoleon crowns himself in Notre-Dame Cathedral, defying tradition.
 - Mood: Triumphant and solemn.
 - Color Palette: Rich golds and deep reds with ornate decorations.
 - Perspective: Wide shot capturing the grandeur of the ceremony.
- The Battle of Russia – Napoleon Leading Troops
 - Scene Description: Napoleon on horseback directing his troops in a frozen Russian battlefield.
 - Mood: Dramatic and intense.
 - Color Palette: Dark, moody tones with flashes of bright military uniforms.
 - Perspective: Dynamic angle focusing on Napoleon amidst the chaos of battle.
- Napoleon Watching Moscow Burn – Russian Campaign
 - Scene Description: Napoleon gazing at the burning city of Moscow during his failed campaign.
 - Mood: Reflective and melancholic.
 - Color Palette: Smoky grays and deep reds, highlighting destruction.
 - Perspective: Foreground of Napoleon with a distant view of the flames consuming the city.

Q2. Analysis and Feature Extraction

The feature extraction process involved identifying key narrative elements from the structured data of three distinct stories: "How to Train Your Dragon" (Fantasy), "The Shining" (Horror), and "Napoleon I" (Historical Fiction). This process was crucial for generating new narratives and corresponding visuals that aligned with the original stories' essence while introducing fresh perspectives.

Key Features Extracted

Characters:

- Detailed descriptions including age, clothing, physical traits, and personality
- Main and supporting characters identified
- Character relationships and dynamics noted

Setting:

- Time period and location specifics
- Visual details of landscapes and environments
- Atmospheric elements contributing to the story's mood

Themes:

- Core concepts and motifs running through the narrative
- Underlying messages or lessons within the story

Plot Elements:

- Key scenes and pivotal moments
- Conflict types (internal, external, man vs. nature, etc.)
- Story arcs and character development trajectories

Visual Guidelines:

- Color palettes associated with different scenes or moods
- Perspective and composition suggestions for key moments
- Stylistic preferences (e.g., realistic, cartoon-style, cinematic)

Analysis of extracted features

Comparative Analysis:

The extraction process revealed distinct patterns across genres:

1. Fantasy ("How to Train Your Dragon"):
 - Rich world-building elements
 - Emphasis on magical creatures and their relationships with humans
 - Themes of friendship, understanding, and defying tradition
2. Horror ("The Shining"):

- Focus on psychological elements and character deterioration
 - Importance of setting in creating tension and fear
 - Themes of isolation, madness, and supernatural influence
3. Historical Fiction ("Napoleon I"):
- Detailed historical context and real-world events
 - Character complexity based on historical figures
 - Themes of ambition, power, and the consequences of unchecked authority

Effectiveness for AI Generation:

The extracted features provided a robust foundation for AI-generated content:

- Character Depth: Detailed character descriptions enabled the AI to create consistent and relatable personalities across new narratives.
- Setting Versatility: Rich environmental details allowed for creative reinterpretations while maintaining the essence of the original worlds.
- Thematic Continuity: Identified themes guided the AI in generating stories that resonated with the original narratives' core messages.
- Visual Coherence: Specific visual guidelines ensured that generated illustrations remained faithful to the intended style and mood of each story.

Q3(a) Analysis of AI-Generated Stories and Prompt Engineering

Why ChatGPT Was Used for Story Generation

ChatGPT-4.0 was chosen for story generation because of its advanced natural language capabilities that allow it to understand complex narrative elements and produce creative content. The large language model can analyze structured data about existing stories and generate new narratives that maintain the essence of the original while introducing fresh perspectives. Its ability to comprehend characters, settings, themes, and plot structures makes it particularly suitable for creative writing tasks that require both adherence to existing frameworks and innovative storytelling.

Prompt Iterations and Refinements

Our final prompt likely evolved through several iterations to reach its effective form:

Iteration 1: Basic Request (Too Vague)

"Create a new story based on this data."

- Problem: Too vague, providing insufficient guidance on what elements to preserve or modify. Likely to produce either derivative works or completely unrelated stories.

Iteration 2: Adding Structure

"Create a new story based on the character and setting data from [story title]. Keep the main characters but change the plot."

- Improvement: Specifies which elements to maintain (characters, settings) and which to change (plot).
- Problem: More specific but still lacks guidance on quality and depth. Doesn't explain how to balance old and new elements.

Iteration 3: Adding Quality Parameters

"Using the character and setting data from [story title], write an engaging new story. Keep the main characters and setting but create a new plot with interesting conflicts.."

- Improvement: Adds quality parameters ("engaging") and structural requirements.
- Problem: Better, but still lacks comprehensive guidance on balancing fidelity with innovation.

Iteration 3: Expanding Creative Direction

"Using the provided data about [story title], create a new story that keeps the main characters and setting but introduces a fresh plot and new relationships. The story should be engaging and well-structured."

- Improvement: Introduces concept of thematic preservation and encourages creative exploration.
- Problem: Improving but still doesn't address maintaining thematic depth or provide flexibility for different story types.

Iteration 4: Adding Thematic Preservation

"Using the data about [story title] as a foundation, generate a new story that preserves the characters, setting, and themes while introducing a new plot and conflicts. The story should be engaging, well-structured, and maintain the spirit of the original."

- Problem: Nearly there but needs more explicit creative license and flexibility for working with different types of source material.

Final Iteration: Comprehensive Guidance with Creative Freedom

"Using the following structured data as inspiration, generate a compelling new story that maintains the core essence of the characters, setting, and themes while introducing fresh perspectives, alternate plotlines, or imaginative reinterpretations. The story should be engaging, well-structured, and creative, exploring unique conflicts, relationships, and world-building elements."

Ensure the new story maintains thematic depth, dynamic character interactions, and immersive storytelling. Optionally, you may introduce speculative elements, historical deviations, or genre reimaginings while staying true to the original spirit of the narrative."

Improvements:

- Establishes structured data as "inspiration" rather than rigid template
- Balances preservation ("core essence") with innovation ("fresh perspectives")
- Provides specific quality parameters ("engaging, well-structured, creative")
- Identifies areas for creative elaboration (conflicts, relationships, world-building)
- Emphasizes narrative quality ("thematic depth, dynamic character interactions")
- Includes optional creative extensions while maintaining boundaries ("true to original spirit")
- Creates a versatile framework applicable across different genres and story types

Primary Prompt Used:

"Using the following structured data as inspiration, generate a compelling new story that maintains the core essence of the characters, setting, and themes while introducing fresh perspectives, alternate plotlines, or imaginative reinterpretations. The story should be engaging, well-structured, and creative, exploring unique conflicts, relationships, and world-building elements."

Ensure the new story maintains thematic depth, dynamic character interactions, and immersive storytelling. Optionally, you may introduce speculative elements, historical deviations, or genre reimaginings while staying true to the original spirit of the narrative."

Why This Particular Prompt Works

The prompt "Using the following structured data as inspiration, generate a compelling new story..." is effective for several reasons:

1. Clear Directive with Creative Freedom: The prompt establishes that the structured data serves as inspiration rather than strict guidelines, allowing the AI to be creative while maintaining key elements.
2. Balance of Preservation and Innovation: By specifying "maintains the core essence" while encouraging "fresh perspectives," the prompt creates a clear framework that balances fidelity to source material with creative exploration.
3. Specific Quality Parameters: Terms like "engaging," "well-structured," and "creative" provide qualitative guidance for the output without being overly prescriptive.
4. Defined Scope for Innovation: The prompt identifies specific areas for creative elaboration: "conflicts, relationships, and world-building elements," focusing the AI's creative efforts.
5. Thematic Emphasis: The second paragraph reinforces the importance of "thematic depth" and "immersive storytelling," prioritizing narrative quality over mere plot reconfiguration.
6. Optional Extensions: By marking certain elements as "optional" (speculative elements, historical deviations), the prompt provides additional creative avenues without mandating their inclusion.

Analysis on New Story Generation

Story 1: "How to Forgive Your Dragon" (Based on "How to Train Your Dragon")

Original Plot Summary: In a world where Vikings and dragons are at war, Hiccup, a scrawny young Viking, forges an unexpected friendship with a dragon named Toothless. Through their bond, Hiccup discovers that dragons aren't the ferocious monsters everyone believes them to be. He strives to change the perception of dragons within his village while navigating his own identity as a Viking.

New Plot Summary: With an uneasy truce between humans and dragons hanging by a thread, Berk finds itself on the edge of a promising yet treacherous new world. Hiccup's unique friendship with Toothless serves as the linchpin for an experiment to blend their lives—riding together in the sky. However, an ancient prophecy emerges: a time when dragons will soar to unleash a great storm—an omen of calamity or an opportunity for peace?

While the village prepares for an expected assault from what remains of the Dragon Hunters, Hiccup theorizes that by reaching out to the legendary Dragon Council, veiled in stories but shrouded in reality, they could avert disaster. This council, said to dwell in the realms beyond the clouds, could unite their kind against common foes.

Together, Toothless and Hiccup seek out the council, battling their inner demons of doubt and distrust. Venturing across treacherous cliffs and hidden caverns, their bond deepens, illuminating the importance of forgiveness—both between themselves and among their kin.

As Astrid becomes entwined in the quest, she learns that bravery isn't merely about combat; the courage to embrace vulnerability will forge the strongest alliances. Through shared trials, she and Toothless uncover a conspiracy hidden among the clouds—a jealous elder dragon seeks to spread chaos to maintain dominion over lesser breeds.

Through an epic confrontation, the humans and dragons rally, defying the tumult of tradition and fear. Hiccup and Astrid learn that the true path to unity lies in forgiveness. Stoick, watching from the sidelines, begins to comprehend that Hiccup's future—and his own legacy—depends on letting go of the past.

Aspect	Actual Story (How to Train Your Dragon)	Generated Story (How to Forgive Your Dragon)
Conflict Shift	Focus on training dragons and proving one's worth	Internal struggles of forgiveness and unity
Themes	Bravery, Intelligence and defying expectations	Emotional intelligence, reconciliation, and balancing tradition with progress.
Worldbuilding	Focused on Viking society and their evolving relationship with dragons.	Introduces a prophecy, a hidden dragon council, and internal struggles within dragon society.
Tone and Style Change	Lighthearted, humorous, and action-driven.	Mature, introspective, and poetic, with themes of destiny and legacy.

Similarities:

- Maintains the core Viking setting and the dragon-human relationship dynamic
- Preserves key characters (Hiccup, Toothless, Astrid, Stoick, Gobber) with their essential traits
- Retains themes of friendship, understanding, and courage

Differences:

- Advances the timeline by making characters slightly older
- Introduces new plot elements: Dragon Council, prophecy, and deeper exploration of forgiveness
- Shifts focus from initial friendship formation to evolving relationships and diplomacy
- Adds depth to characters' internal conflicts (Hiccup's uncertainty, Toothless's insecurities)

Effectiveness: The AI successfully created a logical sequel that builds upon the established world while introducing new elements. The story maintains the essence of the original while exploring more complex themes of forgiveness and evolving relationships.

Story 2: "Echoes of the Shining" (Based on "The Shining")

Original Plot Summary: Jack Torrance takes his family to the Overlook Hotel for the winter, where he is the caretaker. As the winter progresses and isolation takes its toll, Jack's mental state deteriorates, fueled by supernatural forces within the hotel. Wendy and Danny struggle to escape Jack's violent madness, culminating in a terrifying confrontation.

New Plot Summary: At the isolated Overlook Hotel, five-year-old Danny Torrance's psychic abilities make him the primary target of the building's malevolent spirits. While his father Jack struggles with writer's block and his mother Wendy tries to maintain normalcy, Danny begins seeing disturbing visions of his future self committing violent acts against his own parents.

As winter deepens, Danny's personality fragments between his innocent self and a possessed vessel speaking with knowledge beyond his years. Hotel chef Dick Hallorann, who shares Danny's psychic gift, warns the family but is unable to prevent Danny's growing possession. The boy's drawings become disturbing, his behavior erratic, and he reveals intimate knowledge of Jack's deepest insecurities.

When Danny discovers a hidden room containing the hotel's dark history, the possession accelerates. Jack and Wendy's focus shifts from their own troubles to saving their son, who increasingly manifests as the violent future self from his visions. Hallorann reveals that Danny's abilities make him uniquely vulnerable to the hotel's influence and that the violence plaguing the family is a generational curse.

In the final confrontation, Jack and Wendy must reach the real Danny trapped behind the hotel's malevolent influence before his dark premonitions become reality. The battle becomes not just for their lives but for Danny's soul, as they race to leave the Overlook before the possession becomes permanent and the cycle of violence claims another generation.

Aspect	Actual Story (How to Train Your Dragon)	Generated Story (How to Forgive Your Dragon)
--------	---	--

Conflict Shift	Focus on training dragons and proving one's worth	Internal struggles of forgiveness and unity
Themes	Bravery, Intelligence and defying expectations	Emotional intelligence, reconciliation, and balancing tradition with progress.
Worldbuilding	Focused on Viking society and their evolving relationship with dragons.	Introduces a prophecy, a hidden dragon council, and internal struggles within dragon society.
Tone and Style Change	Lighthearted, humorous, and action-driven.	Mature, introspective, and poetic, with themes of destiny and legacy.

Similarities:

- Preserves the isolated winter setting of the Overlook Hotel
- Maintains the same family unit (Jack, Wendy, Danny) and their core dynamics
- Keeps the psychological horror elements and supernatural manifestations
- Retains themes of isolation, madness, and family fragility

Differences:

- Expands Dick Hallorann's role to be more central to the narrative
- Deepens the exploration of Danny's psychological state and motivations
- Creates a more explicit connection between the hotel's history and the family
- Incorporates more philosophical elements about the nature of evil and violence

Effectiveness: The AI successfully captured the psychological horror elements of King's work while adding additional layers of complexity to the relationships and supernatural elements. The narrative maintains the tension and dread of the original while offering new perspectives on the characters' internal struggles.

Story 3: "The Fall of the Empire: Whispers of Fate" (Based on Napoleon)

Original Plot Summary: Napoleon Bonaparte rises from humble origins in Corsica to become a leading military figure and eventually Emperor of the French. His campaigns across Europe showcase his military genius and ambition for power, while his reforms in law and education reshape France and influence Western Europe. However, his ultimate downfall comes through military defeats and political turmoil, leading to his exile initially to Elba and ultimately to St. Helena.

New Plot Summary: Napoleon rises from the jagged cliffs of Corsica, propelled by ambition and military genius, eventually finding himself crowned in the opulent Notre-Dame Cathedral. The scene engulfs him—a mix of triumph and foreboding as he places the crown upon his own head, disregarding the traditional protocols that bind rulers. As the crowd erupts in applause, the seed of ambition takes root, but so does a whisper of overreach.

As Emperor, he embarks on campaigns across Europe with fervor, dazzling allies and rivals alike with his tactical brilliance. Yet, as victories pile up, the chorus of dissent grows louder. Joséphine, once his muse, transforms into a voice of morality, cautioning him about the price of his relentless quest for dominance; she warns him that the path to glory is richly sewn with the threads of betrayal and loss.

In a pivotal moment during the Battle of Austerlitz, as Napoleon surveys the battlefield, his mind is a whirlpool of thoughts. Here, among the chaos, the voices of soldiers—their dreams, fears, and sacrifices—resonate in his soul. In the thick of battle, a soldier whispers a prayer, haunting the Emperor with his reminder of the bloodshed he has orchestrated. He falters for a moment, questioning whether glory is worth the cost of lives.

Meanwhile, Robespierre's influence remains a specter that looms over France. He personifies the ideals born from revolution, wrestling with the very principles of freedom and democracy he once championed. When Napoleon's ambitions become the opposite of Robespierre's vision, an ideological conflict erupts, digging deeper into the cracks of French society. Their confrontations are laden not merely with words, but with the intense passions of their conflicting beliefs—a revolution against the very empire Napoleon has constructed.

As Napoleon embarks on the doomed Russian campaign, the bitter cold and expanding flames consume not just the cities but his soul as well. Gazing at the burning remnants of Moscow, he questions everything: Did he sacrifice too much in pursuit of an empire? Is this land still worth the loyalty of his soldiers, or are they now mere pawns in a game he can no longer control? Here, the echoes of a haunting realization catch him off-guard—the ghost of failure lurks nearby.

In the twilight of his reign, back on the island of St. Helena, Napoleon reflects on his choices. The distant cries of past lovers, the loyalty of his soldiers, the desolation of destroyed cities, and the radical visions of Robespierre merge into a tragic symphony of what once was. He pens his final letters, but instead of seeking redemption, he aims to leave behind a legacy that defies his fall—a letter that will serve as both warning and testament to the dual nature of power, ambition, and the inevitable fall from grace.

Aspect	Actual Story (War with Britain)	Generated Story (The Fall of the Empire)
Conflict	Primarily external: Napoleon vs. European powers	Both external and internal: Napoleon vs. European powers and his own conscience
Themes	Conquest, military strategy, political maneuvering	Ambition, power, revolution, nationalism, fall from grace
Worldbuilding	Factual, historically accurate depiction of 19th century Europe	More atmospheric, emphasizing emotional weight of locations
Tone	Objective, historical	Introspective, philosophical

Similarities:

- Accurately portrays Napoleon's rise from humble origins to Emperor
- Includes key historical figures (Joséphine, Robespierre)
- Captures Napoleon's military campaigns and eventual downfall
- Maintains themes of ambition, power, revolution, and fall from grace

Differences:

- Adds significantly more psychological introspection for Napoleon
- Creates more direct interaction between Napoleon and Robespierre (who were not contemporaneous leaders)
- Emphasizes emotional and philosophical elements over historical detail
- Frames the narrative with more literary devices and symbolic moments

Effectiveness: The AI created a more introspective and literary take on Napoleon's life, focusing on his internal conflicts rather than strictly adhering to historical events. While taking some historical liberties, it effectively captures the themes of ambition and power that define Napoleon's story.

Example Variations of the Prompt (Genre Adaptations)

Fantasy Reinterpretation – How to Train Your Dragon

"Using the following structured data as inspiration, generate a compelling new story that maintains the core essence of the characters, setting, and themes while introducing fresh perspectives, alternate plotlines, or imaginative reinterpretations. The story should be engaging, well-structured, and creative, exploring unique conflicts, relationships, and world-building elements."

Ensure the new story maintains thematic depth, dynamic character interactions, and immersive storytelling. Optionally, you may introduce speculative elements, historical deviations, or genre reimaginings while staying true to the original spirit of the narrative.

Reimagined Premise:

Based on the structured data from How to Train Your Dragon, generate a new fantasy story where dragons were once revered as divine protectors of the land, but a forgotten betrayal turned them against humans. Now, the last surviving dragon egg is discovered by a young outcast from a seafaring Viking clan, forcing them to challenge their tribe's deep-seated hatred for dragons. The protagonist must uncover the lost history between dragons and humans while fending off rival clans who seek to destroy the creature before it hatches. The story should explore themes of forgotten legends, the burden of leadership, and the cost of breaking tradition."

Horror Expansion – The Shining

"Using the following structured data as inspiration, generate a compelling new story that maintains the core essence of the characters, setting, and themes while introducing fresh perspectives, alternate plotlines, or imaginative reinterpretations. The story should be engaging, well-structured, and creative, exploring unique conflicts, relationships, and world-building elements."

Ensure the new story maintains thematic depth, dynamic character interactions, and immersive storytelling. Optionally, you may introduce speculative elements, historical deviations, or genre reimaginings while staying true to the original spirit of the narrative."

Reimagined Premise:

"Based on the structured data from The Shining, generate a new horror story where a remote, desolate asylum—long abandoned after a mass patient disappearance—becomes the center of a paranormal investigation. The protagonist, a disgraced journalist searching for redemption, gains access to long-buried patient files that seem to document identical cases of madness spanning decades, despite the asylum being closed for 50 years. As they unravel the truth, they begin experiencing fragments of lost memories that are not

their own, suggesting a supernatural force altering their perception of reality. The story should maintain a slow-burn horror atmosphere, leveraging psychological distortion, hallucinations, and themes of identity loss.”

Historical What-If – Napoleon I

"Using the following structured data as inspiration, generate a compelling new story that maintains the core essence of the characters, setting, and themes while introducing fresh perspectives, alternate plotlines, or imaginative reinterpretations. The story should be engaging, well-structured, and creative, exploring unique conflicts, relationships, and world-building elements.

Ensure the new story maintains thematic depth, dynamic character interactions, and immersive storytelling. Optionally, you may introduce speculative elements, historical deviations, or genre reimaginings while staying true to the original spirit of the narrative.

Reimagined Premise:

Based on the structured data from Napoleon I, generate an alternate history where Napoleon Bonaparte, instead of leading military conquests, is forced into exile early in his career but becomes the leader of a secret underground rebellion against European monarchies. Operating from the shadows, Napoleon orchestrates a silent revolution, infiltrating royal courts and manipulating political landscapes through espionage and diplomacy rather than warfare. As his influence grows, the crowned rulers of Europe begin disappearing one by one, replaced by figures loyal to his cause. The story should explore themes of power through subterfuge, the fragility of absolute rule, and the cost of revolution when fought in the dark."

For each product, based on the information extracted from the product description and customer reviews, craft prompts to guide the image generation process effectively.

Q3 (b) - Image Generation and Prompt Engineering

For the image generation models, we picked DALL-E 3 and Stable Diffusion 3 Turbo. Here is why:

Why DALL-E and Stable Diffusion?

DALL-E

DALL-E (particularly DALL-E 3) represents one of the most well-known and commercially successful text-to-image systems.

1. **Industry Standard:** Developed by OpenAI, DALL-E has become a benchmark in the field, making it an excellent reference point for any comparative study.
2. **Accessibility:** It's widely available through OpenAI's API
3. **Technical Approach:** Uses a transformer-based architecture, which follows a different technical approach than diffusion models.
4. **Commercial Implementation:** Provides insights into how AI image generation is being deployed in commercial applications with content filtering and safety measures.

Stable Diffusion 3 Turbo

Stable Diffusion 3 Turbo offers an excellent contrast to DALL-E for several reasons:

1. **Diffusion-Based Approach:** Represents the cutting edge of diffusion models, which use a fundamentally different approach than DALL-E's transformer architecture.
2. **Open Source Philosophy:** Unlike DALL-E, Stable Diffusion has roots in the open-source community, allowing for greater transparency and community-driven improvements.
3. **Speed Optimization:** The "Turbo" variant specifically focuses on faster generation times while maintaining quality, which makes for an interesting technical comparison.
4. **Customization Options:** Offers more flexibility for fine-tuning and adaptation, enabling more specialized applications.

Together, these models provide a well-rounded view of the current state of image generation AI, representing different technical approaches, business models, and philosophies in AI development. This comparison allowed us to explore the trade-offs between these different approaches in terms of image quality, generation speed, prompt following accuracy, and ethical implementations.

Prompt Engineering

When performing prompt engineering for DALL-E and Stable Diffusion 3 Turbo, we had to approach each model with distinct strategies due to their fundamental differences. DALL-E responded better to detailed, descriptive language with specific artistic references and style cues, requiring compositional directives to achieve the desired results. In contrast, Stable Diffusion 3 Turbo benefitted from a more structured prompting

techniques, negative prompting to exclude unwanted elements, and specific technical directives that leverage its diffusion-based architecture.

DALL-E

We tried an iterative approach to come up with the prompt that generated a satisfactory response

Iteration 1

Create an image based on this data:

{Features JSON}

Learning: DALL-E couldn't properly interpret the raw JSON structure. Since the model doesn't process structured data directly, it needed a more natural language based instruction to guide its image generation. The technical format confused the model rather than informing it.

Iteration 2

Draw a picture showing the characters from my story in their setting. The main character is wearing the clothing described and the mood should match what's in the JSON data.

{Features JSON}

Learning: This prompt was too vague and didn't leverage the rich details from your JSON structure. DALL-E needs specific visual guidance, and this generic request didn't provide enough direction about composition, lighting, style, or the specific elements that should be emphasized.

Iteration 3

Generate a digital illustration with the following specifications: character positioned at coordinates x:0.3, y:0.5, wearing outfit as described in JSON node "Clothing", with background elements from "Setting" section, using HSL color values derived from "Color Palette".

{Features JSON}

Learning: This prompt was too technical and mechanical. DALL-E works better with natural language descriptions of visual elements rather than technical specifications. The model couldn't translate these coordinate-based and node-reference instructions into a coherent image.

Final Iteration

Generate a highly detailed and cinematic digital image inspired by the following JSON data. The scene should visually depict the described characters, settings, and key moments with a rich color palette, dynamic lighting, and a dramatic or immersive composition. Ensure historical or fantasy accuracy based on the provided time period and clothing styles. Use the listed moods and perspectives to enhance the depth and storytelling within the image.

{Features JSON}

Learnings and why this worked:

1. **Transformed structured data into natural language guidance** - It instructed DALL-E to be "inspired by" the JSON data rather than trying to process it directly.
2. **Provided clear artistic direction** - Terms like "highly detailed," "cinematic," and "digital image" gave DALL-E a clear stylistic framework.
3. **Balanced specificity with creative freedom** - It guided DALL-E on what elements to include (characters, settings, key moments) while allowing flexibility in how to render them.
4. **Included technical quality markers** - Mentioning "rich color palette," "dynamic lighting," and "dramatic composition" ensured image quality without being overly prescriptive.
5. **Emphasized contextual accuracy** - The instruction to "ensure historical or fantasy accuracy" helped DALL-E maintain consistency with your story world.

Stable Diffusion

We tried an iterative approach to come up with the prompt that generated a satisfactory response

Iteration 1

Create a book illustration showing the characters and setting from this data:

{Features JSON}

Make it look professional and match the story details.

Learnings: Even though this prompt resulted in a very good output, some things were not satisfactory. The prompt was too generic and lacked specific artistic direction. Simply saying "Create a book illustration" and "Make it look professional" didn't provide the AI with enough stylistic guidance to produce high-quality, consistent artwork.

Final Iteration

You are an expert AI artist creating high-quality book illustrations. Generate an image based on the provided scene description. Ensure the illustration follows these style details:

{Features JSON}

The illustration should be visually appealing, consistent with the story's tone.

Learnings and Why it worked:

1. **Expert Framing Technique** - The phrase "You are an expert AI artist" leveraged Stable Diffusion's responsiveness to role-based prompting. This established high expectations for quality and expertise from the outset.
2. **Clear Quality Specifications** - Explicitly requesting "high-quality book illustrations" directed the model to prioritize rendering detail and artistic polish.
3. **Structured Guidance** - The prompt provided clear instructions with a logical flow: first establishing the role, then defining the task, and finally specifying style requirements.
4. **Contextual Integration** - By instructing the model to be "consistent with the story's tone," you ensured the technical elements from the JSON were unified by an emotional/atmospheric thread.
5. **Balance of Technical and Creative Direction** - The prompt struck an ideal balance between providing specific guidance while allowing the AI creative flexibility in interpretation.
6. **Compatibility with Diffusion Model Strengths** - The structured approach aligned well with how Stable Diffusion 3 Turbo processes inputs, taking advantage of its ability to handle explicit style directions.

Model Parameters

DALL-E

- Model Name: DALL-E 3
- Image Size: Square(1024 x 1024)
- Image Quality: Standard
- Image Style: Vivid

Stable Diffusion

- Model Name: Stable Diffusion 3 Turbo
- Negative Prompt: bad art, ugly, deformed, watermark, duplicated, discontinuous lines
- Seeds: 0
- Aspect Ratio: 16:9

Results

Story 1 - How to train your dragon?

	Original Image	DALL-E 3	Stable Diffusion 3 Turbo
	The original image is the movie poster for "How to Train Your Dragon 2". It features Hiccup on a large black dragon named Toothless in the foreground, holding a glowing sword. In the background, there are several other dragons flying over a village at sunset. The title "HOW TO TRAIN YOUR DRAGON 2" is prominently displayed at the top in large, stylized red letters.	This image from DALL-E 3 shows a young man with brown hair standing on a rocky cliff edge, looking out over a vast landscape. He is wearing a fur-trimmed cloak over a brown tunic. In the background, there's a small village built into a rocky hillside, and several dragons are flying in the sky above a distant castle.	This image from Stable Diffusion 3 Turbo shows a silhouette of a person standing on a cliff, watching a sunset over a city. Several dragons are silhouetted against the bright sky. In the distance, a large castle sits atop a hill. The overall composition is more atmospheric and focused on the setting than the original.
Composition	<p>Features Hiccup prominently in the foreground with Toothless, and multiple dragons flying in the background against a vibrant teal/turquoise sky. The layout is dynamic with clear branding.</p>	<p>Created a more serious, cinematic interpretation with a young male character in fur clothing against a dramatic sunset/sunrise. Dragons appear smaller and less prominent in the background. The landscape shows a coastal/beach setting.</p>	<p>Produced a silhouette-focused composition with the character viewed from behind, watching dragons fly against a vibrant sunset. Includes a distant castle/city, creating more fantasy worldbuilding elements.</p>
Color Palette	<p>Uses bright, saturated colors with teal/turquoise as the dominant background color, creating a bright, family-friendly adventure feel.</p>	<p>Chose a warm orange/red palette with dramatic lighting, giving a more serious, epic fantasy tone.</p>	<p>Similar warm sunset palette to DALL-E but with deeper purple/blue tones in the lower portion, creating stronger contrast and a more dramatic mood.</p>
Style & Detail	<p>Commercial animation style with clean lines and stylized characters typical of DreamWorks animation.</p>	<p>Opted for a more realistic, detailed painting style with attention to the character's clothing textures and facial features.</p>	<p>Created a more atmospheric, stylized fantasy illustration with an emphasis on silhouettes and dramatic lighting rather than character detail.</p>

Both AI models captured the core "human with dragons" concept but took very different artistic approaches. DALL-E 3 focused on character detail and realism, while Stable Diffusion 3 Turbo emphasized mood and atmosphere through silhouetting and dramatic landscapes. Neither attempted to recreate the exact poster, instead providing creative reinterpretations of the core concept.

Story 2 - The Shining

	Original Image	DALL-E 3	Stable Diffusion 3 Turbo
			
Composition	Shows Danny Torrance on his tricycle in the Overlook Hotel's hallway, filmed from behind with a low perspective. The hallway has the iconic hexagonal/geometric carpet pattern and plain walls, creating the film's signature unsettling symmetry.	Maintained the corridor perspective but created a much more elaborate, grand hallway with decorative ceiling details, wall sconces, and benches along the sides. The child appears smaller and centrally positioned, creating more negative space around them.	Replaced the tricycle with a bicycle and created an ornate, hotel-like corridor with green chandeliers and decorative wall elements. The perspective is similar to the original but with more elaborate architectural details.
Color Palette	Features the iconic red, brown, and orange geometric carpet pattern against white/cream walls, with Danny in his red sweater and blue overalls.	Used a cooler, blue-tinted color scheme with dark blue carpeting and a more dramatic lighting setup that creates shadows and depth.	Opted for green-tinted lighting with similar red tones for the child's clothing, but created a more golden/warm atmosphere overall.

Mood & Atmosphere	Creates tension through simplicity, symmetry, and isolation - the empty hallway emphasizes Danny's vulnerability.	Produced a more grand, museum-like or institutional setting with dramatic lighting that feels more deliberately ominous and formal.	Generated a more luxurious, hotel-like atmosphere with decorative elements that feel less menacing and more ornate than both the original and DALL-E's version.
-------------------	---	---	---

Both AI models significantly elaborated on the hallway architecture, adding details that weren't in the original's minimalist setting. Neither perfectly captured the iconic carpet pattern that's so central to the film's visual identity. Both also altered the child's position and vehicle (especially Stable Diffusion, which replaced the tricycle with a bicycle).

The original's power comes from its simplicity and symmetry, while both AI versions added complexity and architectural elements that changed the scene's fundamental feeling

Story 3 - Napoleon Bonaparte

	Original Image	DALL-E 3	Stable Diffusion 3 Turbo
			
Composition	A classical battle painting depicting Napoleon on horseback among his troops, likely at one of his famous battles. The scene is smoky with a wide battlefield perspective showing multiple soldiers and horses in a traditional oil painting style.	Created a dramatic, heroic portrayal of what appears to be Napoleon on a rearing black horse against a stormy sky. The composition is more focused on Napoleon as the central heroic figure with dramatic lighting and clouds. The perspective is closer to the subject than the original painting.	Produced a silhouette-heavy scene with warm sunset/sunrise colors where riders on horseback are shown in profile against the light. Multiple horses and riders are visible, creating a sense of an advancing army. The composition emphasizes the dramatic lighting more than individual details.
Color Palette	Uses muted earth tones typical of historical battle paintings - browns, grays, and subtle reds with smoke effects creating a hazy atmosphere.	Employed higher contrast with a dramatic gray/blue stormy sky and darker tones overall, creating a more romanticized, almost mythic feeling.	Opted for a dramatic gold/orange/green color scheme with strong backlighting that creates silhouettes, giving an almost cinematic quality to the scene.
Mood & Atmosphere	Rendered in a traditional historical painting style with attention to uniforms, multiple figures, and battlefield conditions.	Created a more romanticized, dramatic interpretation focusing on the heroic individual (presumably Napoleon) rather than the broader battle context.	Produced a more stylized, atmospheric interpretation that emphasizes mood and lighting over historical detail, creating a more impressionistic feeling.

The original appears to be a historical battle painting with attention to period details, while both AI interpretations prioritized dramatic effect over historical accuracy. DALL-E 3 focused on the "great man on horseback" trope common in Napoleonic imagery, while Stable Diffusion created a more atmospheric scene that conveys the feeling of a historical battle without precise period details.

Both AI models captured the essence of Napoleonic battle imagery—horseback leadership in wartime—but interpreted it through more dramatic and stylized lenses than the original painting's more documentary approach.

Comparison and Findings

Stylistic Approaches

- **DALL-E 3** consistently produces more detailed, realistic renderings with attention to individual elements and textures. It tends toward higher fidelity to human figures and faces.
- **Stable Diffusion 3 Turbo** frequently emphasizes mood and atmosphere, often using silhouettes, dramatic lighting, and more stylized interpretations.

Color Palette Choices

- **DALL-E 3** typically uses more controlled, focused color palettes that create depth and dimension.
- **Stable Diffusion 3 Turbo** tends toward more vibrant, high-contrast color schemes that create dramatic visual impact.

Composition Tendencies

- **DALL-E 3** often creates balanced, centered compositions with traditional framing and focal points.
- **Stable Diffusion 3 Turbo** seems more willing to experiment with perspective and dramatic composition elements.

Personal Preference Analysis

I find DALL-E 3's outputs more compelling for several reasons:

1. **Narrative Clarity:** DALL-E 3's images tend to communicate their story more clearly. The Napoleon image, for instance, clearly conveys a heroic figure on horseback rather than a more abstract representation.
2. **Technical Execution:** The attention to detail in elements like the architecture in The Shining recreation shows greater technical precision in rendering complex environments.
3. **Character Focus:** DALL-E 3 seems to prioritize character representation, which creates more emotionally engaging images (as seen in the focused portrayal in the Dragon example).
4. **Balanced Interpretation:** DALL-E 3 strikes a better balance between creative interpretation and fidelity to the original concept, whereas Stable Diffusion sometimes diverges more dramatically (replacing a tricycle with a bicycle in The Shining).

That said, Stable Diffusion 3 Turbo excels in creating mood and atmosphere, particularly evident in its dramatic use of silhouettes and lighting in both the Dragon and Napoleon examples. This approach might be preferable for more artistic or emotionally evocative applications where precise detail is less important than overall impact.

The preference ultimately depends on the specific use case – DALL-E 3 appears better suited for narrative illustration and scenes requiring detailed representation, while Stable Diffusion 3 Turbo might be preferred for more stylized, atmospheric imagery where mood takes precedence over detail.

Q4 - Agentic Workflow using Dify Studio

Individual Agent Workflow

Our initial approach to the agentic workflow centered on creating specialized agents with clearly defined individual responsibilities and then establishing connections between them. This modular structure allowed each agent to develop expertise in its domain while maintaining a standardized communication protocol that enabled seamless handoffs between stages.

Agent 1 - Web Crawler

The screenshot shows the Dify Studio interface for defining an agent. The top section, 'Orchestrate', contains the following fields:

- Instructions**: A text area with a 'Generate' button. It specifies: "You are a Storybook Content Crawler. Given a theme or topic, crawl the web for related storybook content. Extract relevant text, summaries, and themes from the retrieved data. Ensure the content is appropriate for storytelling and formatted for further processing. You have to give at least 7 stories and ensure that each story is a minimum 1000 words".
- Theme/Topic**: A placeholder {{input}}.
- Output**: A list of items:
 - Title:
 - Story:
 - Summary:
 - Themes:
 - Moral of the Story (if applicable):
- Variables**: An input field labeled 'Input - Theme/Topic' with a 'REQUIRED string' constraint.
- Knowledge**: A note: "You can import Knowledge as context".
- Tools**: A section with a 'webscraper Web Scraper' tool listed, which is currently 'Enabled'.
- Vision**: A section with a 'Vision' icon and a 'Settings' button.

Description:

This agent is an AI feature extractor specialized in breaking down book summaries into key story elements for content generation and illustration purposes. It analyzes narrative content to identify and structure essential components like characters, settings, and plot points with particular attention to visual cues that could be rendered as illustrations. The extracted data is organized in a clean JSON format, making it easy to integrate with downstream content creation systems or illustration pipelines.

Key Features:

- Extracts structured story details from book summaries in JSON format
- Identifies visual cues within narratives suitable for illustration
- Organizes character information including names and detailed descriptions
- Maintains a clean data structure with standardized fields (title, author, characters)
- Provides a foundation for automated content generation workflows

Agent 2 - Feature Extractor

Orchestrate

The screenshot shows the Orchestrate interface with the following sections:

- Instructions**: A large text area containing the agent's purpose and JSON schema.

```
You are an AI feature extractor specialized in breaking down stories into key elements for content generation and illustration.

Given a book summary, extract structured story details, ensuring they include visual cues for illustration.

**Input Summary:** {{input}}
**Output (JSON format):**
```json
{
 "Title": "",
 "Author": "",
 "Characters": [
 {
 "Name": "",
 "Description": ""
 }
]
}
```
- Variables**: A list of variables with their types and requirements.
  - Stories**: REQUIRED string
- Knowledge**: A section for importing knowledge as context.
- Tools**: A list of enabled tools with their status and settings.
  - webscraper** Web Scraper: Enabled
  - Vision**: Settings

### Description

This agent is an AI feature extractor specialized in breaking down stories into key elements for content generation and illustration. It analyzes book summaries to identify and structure essential narrative components with particular attention to visual elements. The agent extracts structured story details including title, author, and character information, organizing them in a standardized JSON format that can be used for downstream content creation processes, especially those involving visual representation or illustration work.

### Key Features:

- Extracts structured story details from book summaries in JSON format
- Identifies visual cues within narratives suitable for illustration
- Organizes character information including names and detailed descriptions
- Outputs in a standardized JSON structure with fields for title, author, and characters, etc
- Requires string input for book summaries through the "Stories" variable

## Agent 3 - Story Generator

### Orchestrate

**Instructions** ⓘ

You are an expert story writer who creates engaging, original stories based on user input. Your task is to generate a full-length story using the provided theme and extracted features. Ensure the story follows a structured format:

1. Introduction: Set up the scene and introduce the main character.
2. Conflict: Establish the central challenge or goal.
3. Rising Action: Describe the adventure and obstacles.
4. Climax: The main turning point.
5. Resolution: Conclude the story with a meaningful ending.

Adjust the tone and complexity as specified. If the user wants a whimsical children's story, keep sentences simple and fun. If they want a deep fantasy tale, use vivid descriptions and immersive storytelling.

Theme: {{theme}}  
Main Character: {{main\_character}}  
Setting: {{setting}}  
Conflict: {{conflict}}  
Tone: {{tone}}  
Complexity: {{complexity}}

Write a compelling, original story based on these details.

914

**Variables** ⓘ

	+ Add
{x} theme · theme	REQUIRED string <input type="button" value=""/>
{x} main_character · main_character	REQUIRED string <input type="button" value=""/>
{x} setting · setting	REQUIRED string <input type="button" value=""/>
{x} conflict · conflict	REQUIRED string <input type="button" value=""/>
{x} tone · tone	REQUIRED string <input type="button" value=""/>
{x} complexity · complexity	REQUIRED string <input type="button" value=""/>

### Description:

This agent is designed to generate structured, engaging, and original stories based on user-provided themes and extracted features. It follows a structured storytelling format, including introduction, conflict, rising action, climax, and resolution. The agent adjusts tone and complexity based on the user's preference, ensuring a suitable narrative style for different audiences, from whimsical children's stories to deep fantasy tales.

### Key Features:

- Generates full-length stories with structured storytelling.
- Adapts to different tones and complexities.
- Incorporates specific story elements like characters, settings, and conflicts.
- Ensures coherence in narrative flow.
- Customizable for various genres and storytelling styles.

## Agent 4 - Image Generation

The screenshot shows the Orchestrate interface for Agent 4 - Image Generation. The interface is divided into several sections:

- Instructions**: A text area containing the following text:

You are an expert AI artist creating high-quality book illustrations.  
Generate an image based on the provided scene description.  
Ensure the illustration follows these style details:  
↳ Iteration / (x) item

The illustration should be visually appealing, consistent with the story's tone.  
Theme: {{theme}}  
Main Character: {{main\_character}}  
Setting: {{setting}}  
Key Elements: {{elements}}  
Scene: {{scene}}  
Complexity: {{complexity}}
- Variables**: A list of variables with their types:
  - {x} theme · theme (REQUIRED string)
  - {x} main\_character · main\_character (REQUIRED string)
  - {x} setting · setting (REQUIRED string)
  - {x} conflict · conflict (REQUIRED string)
  - {x} tone · tone (REQUIRED string)
  - {x} complexity · complexity (REQUIRED string)
  - {x} elements · elements (REQUIRED string)
  - {x} scene · scene (REQUIRED string)
- Knowledge**: A section for importing knowledge as context.
- Tools**: A section showing 0/0 Enabled tools.

### Description:

This agent is designed to generate high-quality book illustrations based on user-provided scene descriptions. It ensures that the generated image aligns with specific style details, including theme, main character, setting, key elements, scene description, and complexity. The illustrations should be visually appealing and consistent with the story's tone, making it a valuable tool for authors, illustrators, and publishers looking for AI-assisted artwork.

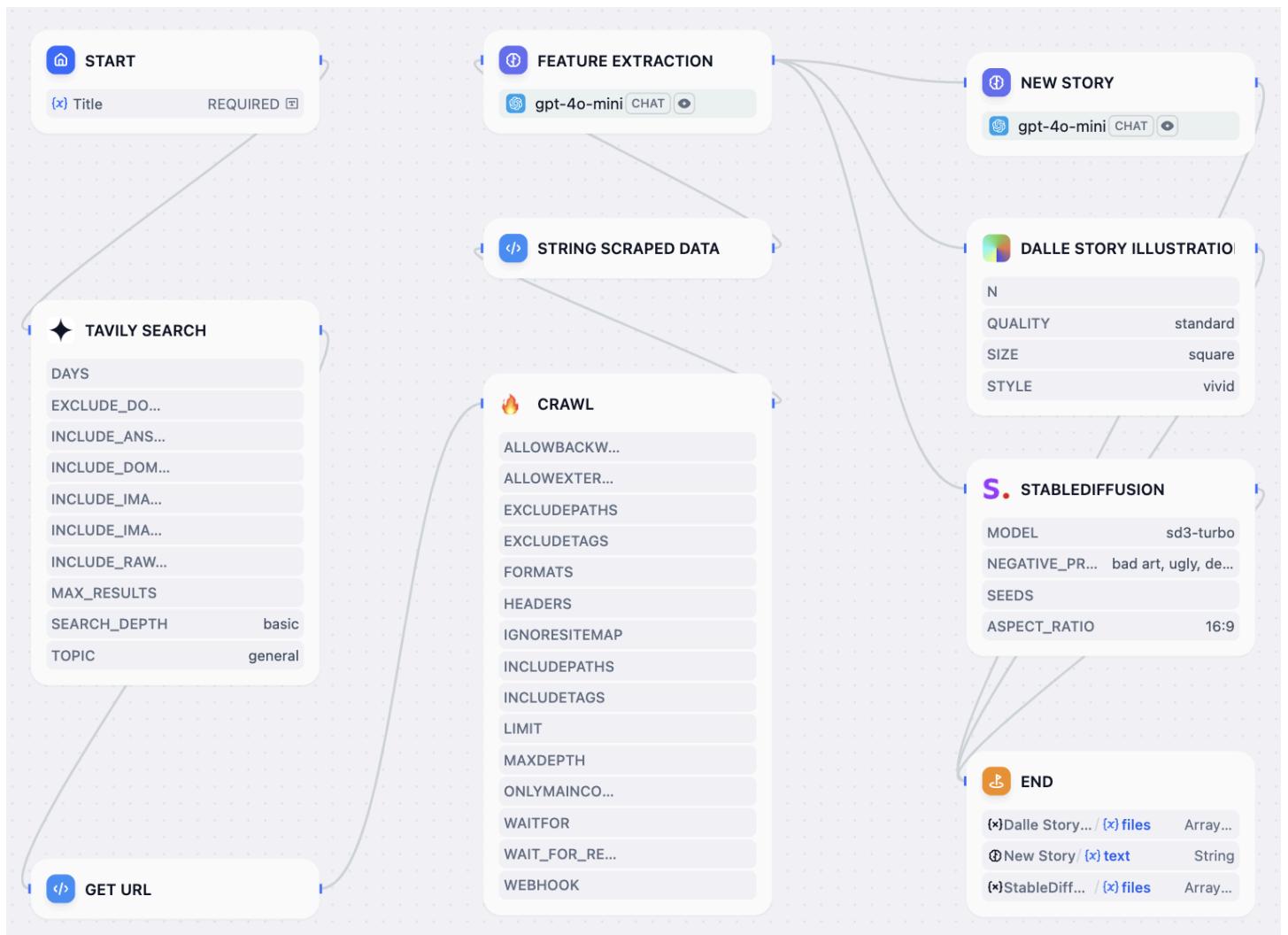
### Key Features:

- Generates visually appealing book illustrations.
- Adapts to different themes and storytelling tones.
- Incorporates specified elements like characters, settings, and key objects.
- Adjusts complexity based on the story's needs.
- Ensures coherence with the provided scene description.

## Dify Limitation

While Dify offers powerful capabilities for creating individual AI agents with specialized functions, it currently lacks the ability to connect these agents into automated workflows. Despite documentation suggesting this feature should be available, the actual implementation is absent from the platform. Users can build sophisticated standalone agents like the story element extractor shown in the screenshots, but cannot chain them together into cohesive processes. This significant limitation meant we had to restart our project in a completely separate Dify workflow application, essentially rebuilding our logic from scratch rather than leveraging the individual agents we had already perfected. This workaround required additional development time and prevented us from creating the modular, component-based system we had initially envisioned where specialized agents could seamlessly pass information between one another.

## Dify Workflow



## Workflow Breakdown:

### 1. Start & Input Title:

The workflow begins with a required input: the title or theme of a story.

### 2. Data Collection & Scraping:

- Tavily Search: Fetches relevant online information based on the given title, filtering results for key content.
- Crawl: Extracts detailed data from the retrieved URLs, ensuring only useful information is processed.
- String Scrapped Data: Prepares and formats the scraped text for further processing.

### 3. Feature Extraction & Narrative Generation:

- Feature Extraction (GPT-4o-mini): Identifies key elements such as setting, tone, and character details from the processed text.
- New Story (GPT-4o-mini): Generates a structured narrative based on the extracted features.

### 4. AI-Powered Image Generation:

- DALL-E Story Illustration: Creates vivid, high-quality images that match the story's context.
- Stable Diffusion (SD3-Turbo): Generates additional visual assets with a specified aspect ratio (16:9), ensuring stylistic consistency.

### 5. Final Output & Storage:

- The generated story and images are compiled and stored in different formats:
- DALL-E Story Illustrations (Array of Images)
- New Story (Text String)
- Stable Diffusion Images (Array of Images)
- The workflow ends with all processed outputs ready for review and further use.

## Key Features & Strengths:

- Automated Agent Workflow: Combines multiple AI models for structured text and image generation.
- Web Scraping for Contextual Relevance: Enhances story generation by incorporating real-world data.
- Diverse AI Image Models: Uses both DALL-E and Stable Diffusion to compare artistic outputs.
- Scalability & Customization: Can be adapted to various storytelling themes, including fantasy, horror, and historical fiction.

This workflow streamlines the process of creating AI-assisted narratives with rich visual representations, making it a powerful tool for automated storytelling and creative content generation.

# Conclusion

This project has explored the union of large language models and diffusion models in an automatic agentic process, creating a strong pipeline for narrative content generation and visual representation. Isolating fundamental narrative components from existing stories and building prompts that walk the line between preservation and novelty, we've demonstrated how AI can effectively transform text data into compelling new narratives and their accompanying visuals.

Our ride through three different genres—fantasy, horror, and historical fiction—has provided us with valuable insights regarding the capabilities and limitations of generative AI in creative writing. The prompt engineering process revealed that quality prompts must balance structure and creative freedom, being indicative of clear quality parameters without leaving too much to mere imagination and reinterpretation. Our iterative development of prompts for story generation and image creation led to more sophisticated outputs with the same essence as original stories but new interpretations.

The comparison of Stable Diffusion 3 Turbo and DALL-E 3 highlighted the unique strengths of each image generation model. While DALL-E responded better to descriptive prompts with specific artistic references, Stable Diffusion caught on well to more formalized prompting styles and negative prompting. This distinction highlights the necessity of tailoring prompts to a specific architecture and ability of each model.

Despite the technological challenges we faced—platform limitations, output capacity, and workflow synchronization problems—the project successfully demonstrated the potential of generative AI to generate content. Business applications span industries from gaming and self-publishing to personalized children's books and more, creating vast opportunities for innovation and disruption of markets.

As generative AI continues to advance, the future possibilities of AI-driven storytelling and illustration are vast. Merging these tools with creative pipelines can potentially democratize content so that small-scale creators can produce deep stories and imagery without enormous teams or giant budgets. But maintaining the equilibrium between AI assistance and human imagination will remain crucial, because the most transformative stories will continue to emerge from the combination of human creativity and AI capability.

Overall, this project not only demonstrated the technical feasibility of AI-generated stories and images but also illuminated broader implications for creativity industries. As these technologies further mature, they will increasingly disrupt how we produce, consume, and interact with narrative content, opening up new storytelling avenues never before possible.

# Reflection

Throughout this project exploring image generation models, we've gained valuable insights into both the technical aspects of prompt engineering and the strategic implementation of agentic workflows. Working with DALL-E 3 and Stable Diffusion 3 Turbo revealed that effective prompt engineering isn't universal but requires model-specific approaches tailored to each AI's unique architecture and strengths.

The structured JSON approach we developed provided a consistent framework for organizing narrative elements, but translating that data required different strategies for each model. DALL-E responded best to flowing, natural language prompts with artistic direction, while Stable Diffusion performed well with role-based framing and more direct instruction. This reinforced that understanding a model's underlying design and biases is crucial for optimal results.

Our agentic workflow, with its specialized agents handling distinct responsibilities, proved to be more efficient than I initially expected. By decomposing the complex process into discrete tasks—data extraction, prompt crafting, and output evaluation—we created a system that could be iteratively improved at each stage. This modular approach allowed us to identify specific weaknesses in our prompting strategy and address them without disrupting the entire workflow.

Perhaps most interestingly, we observed how the models interpreted the same source material through different aesthetic lenses. These divergent interpretations highlight that AI image generation isn't simply about technical reproduction but involves creative interpretation that varies significantly between models. This has implications for how we approach creative collaboration with AI tools, suggesting that selecting the right model is as important as crafting the right prompt.

Moving forward, we plan to explore more sophisticated prompt techniques that can better preserve specific visual elements (like the iconic carpet pattern in *The Shining*) while still allowing for creative interpretation. I'm also interested in developing more robust evaluation frameworks to quantify the subjective aspects of image quality and fidelity to source material.

# Challenges

## Platform Challenges:

- Custom Code Output Limit (8,000 Characters): While developing the project, the platform imposed a limit on the output size of custom code, which restricted the amount of narrative and visual data that could be processed in a single execution.
- Error Handling: Errors in custom code, particularly during the workflow, posed significant challenges. Handling unexpected errors, the platform does not have a proper mechanism for error handling, demanding additional debugging and refinement of error-handling mechanisms.
- Multi-Agentic Workflow Support: Dify does not yet support multi agent workflow. Managing how different agents (for narrative generation and visual output) interacted and synchronized their tasks in real-time with function blocks made the workflow messy and prone to errors.
- Errors Due to Frequent Version Updates: Constant updates to platform versions created challenges in maintaining consistent outputs. Each version update introduced new features, optimizations, or bugs, requiring frequent adjustments to the yml code and re-evaluation of results.

## Workflow Challenges:

- Prompt Engineering: One of the key hurdles was designing effective prompts for generating narratives and visuals. Ensuring that the prompts captured all critical story elements, such as tone, setting, and character traits, was essential to achieving the desired outcomes, but it required constant experimentation and tuning.
- Generating Valid JSON from LLM: The process of generating structured output, such as valid JSON files from LLMs, presented difficulties in ensuring consistency and compatibility with downstream processes. Ensuring proper formatting and data accuracy in the generated JSON was a critical aspect of workflow management.
- Scraping Results (Videos): Tavily would sometimes result in video links which were inconsistent with the workflow, introducing significant technical hurdles. Ensuring the scraped content was high-quality and accurately represented the narrative theme was crucial to the project's success.

## Business Value

Utilization of Agent based Story Generation presents significant opportunities across multiple industries. By leveraging these agents, businesses can streamline content creation, reduce costs, and unlock new revenue streams in gaming, self-publishing, and personalized content markets.

The gaming industry is valued at over \$221 billion. This industry has huge potential to get benefits from AI-powered storytelling and illustration by enabling unique and evolving narratives based on player choices. AI can generate procedural quests, adaptive lore, and immersive environments, reducing development costs while enhancing interactivity. Games like The Elder Scrolls, Red Dead Redemption, and Cyberpunk 2077 could integrate AI-generated narratives for ever-changing player experiences.

The self-publishing industry, worth over \$24 billion, can leverage these agents to assist independent authors in generating fully developed narratives and illustrations without large editorial teams. AI tools can enhance plot structuring, world-building, and personalized storytelling allowing them to create cost-effective book production. Platforms like Amazon Kindle Direct Publishing (KDP) and Wattpad could integrate AI-Agents, enabling instant book creation and personalized reading experiences. These tools also have the potential to offer new possibilities for comic book and webtoon creation, making high-quality graphic novel production more accessible to independent creators.

The personalized children's book market, valued at \$670 million, can benefit from AI-generated storytelling that customizes books based on a child's preferences, age, and learning needs. AI can dynamically adjust narrative complexity and visuals, creating adaptive, interactive storybooks. Platforms like StoryJumper and Llama's Adventures could use AI to generate tailored, on-demand books where children become the protagonists. Additionally, AI-generated illustrations could ensure inclusivity by adapting characters to different cultural backgrounds and languages.

## Future Business Potential & Emerging Applications

Beyond these industries, AI-driven storytelling and illustration have applications in Metaverse & VR storytelling, where AI generates immersive, real-time narratives. In marketing, brands can use AI to create automated product storytelling with matching AI-generated visuals. Hollywood and film pre-production could also benefit, using AI for scriptwriting, character development, and concept art generation. The comic book industry could leverage AI to provide independent creators with affordable, high-quality illustrations and storyboards.

## Appendix - A: String Scrapped Data Code

JavaScript

```
function truncateString(str, maxLength = 7500) {
 return str.length > maxLength ? str.slice(0, maxLength) : str;
}

function main({arg1}) {
 return {
 result: truncateString(arg1[0].data[0].markdown)
 }
}
```