



# From Images to Stories: Exploring Player-Driven Narratives in Games

Edirlei Soares de Lima<sup>1</sup>(✉) , Margot M. E. Neggers<sup>1</sup> ,  
Marco A. Casanova<sup>2</sup> , and Antonio L. Furtado<sup>2</sup>

<sup>1</sup> Academy for AI, Games and Media, Breda University of Applied Sciences,  
Breda, The Netherlands

{soaresdelima.e,neggers.m}@buas.nl

<sup>2</sup> Department of Informatics, PUC-Rio, R. Marquês de São Vicente 225,  
Rio de Janeiro, Brazil

{casanova,furtado}@inf.puc-rio.br

**Abstract.** This paper presents a method for generating player-driven narratives from visual inputs by exploring the visual analysis capabilities of multimodal large language models. By employing Bartle’s taxonomy of player types—Achievers, Explorers, Socializers, and Killers—our method creates stories that are tailored to different player characteristics. We conducted a fourfold experiment using a set of images extracted from a well-known game, generating distinct narratives for each player type that are aligned with the visual elements of the input images and specific player motivations. By adjusting narrative elements to emphasize achievement for Achievers, exploration for Explorers, social connections for Socializers, and competition for Killers, our system produced stories that adhere to established narratology principles while resonating with the characteristics of each player type. This approach can serve as a helping tool for game designers, offering new insights into how players might engage with game worlds through personalized image-driven narratives.

**Keywords:** Story Composition · Image Analysis · Intelligent Agents · Player Types · GPT-4o Vision · Text-to-Image

## 1 Introduction

Images are, first and foremost, *descriptive*. A landscape does more than show a location—it communicates its atmosphere and scale—just as a portrait does more than outline a face; it conveys the subject’s identity and emotional profile. But they can also be *narrative*, displaying an event occurring in a given place and enacted by one or more persons. When a sequence of two or more images is presented, such as step-by-step successive scenes constituting an event, the narrative aspect becomes even more evident. The popular adage that “a picture is worth a thousand words” is not only true in the sense that images clarify what may be difficult to understand in a textual explanation, but also in view

of the rich variety of widely different stories that come to the imagination when contemplating the images. This is particularly relevant in the context of game design, where visual storytelling plays a crucial role in creating immersive and engaging experiences for players.

In literary works, such as novels or poems, the text is the dominant component, whereas images, when inserted as *illustrations*, play the secondary role of inducing readers to mentally visualize the textual content, while serving to embellish the printed pages. Famous illustrators like Gustave Doré have enriched classics such as Dante’s *Divine Comedy* and Perrault’s fairy tales. However, images long preceded writing, with *Narrative Art* tracing back to ancient times. A recent *Nature* article [22] describes a cave painting at Leang Karampuang in Sulawesi, Indonesia, dating back at least 51,200 years, making it “the earliest known surviving example of representational art, and visual storytelling, in the world”.

In games, images such as concept art, character designs, environment sketches, and even in-game screenshots are not just illustrations but are integral to the narrative fabric that shapes the player’s journey. They can help convey complex emotions, hint at backstories, and suggest potential plot developments, all of which enhance player engagement. Recognizing this, we propose a tool named ImageTeller that enables a user to put words to the narrative power of images and image sequences to compose stories inspired by the portrayed scenes.

To further explore the potential of ImageTeller in game design, we present a new approach to narrative generation by incorporating a genre framework based on player archetypes. Drawing from Bartle’s taxonomy [1], which identifies distinct player types—Achievers, Explorers, Socializers, and Killers—our proposed method leads to plots that reflect different player motivations and preferences. By using visual inputs such as game concept art or screenshots, designers can utilize ImageTeller to create stories that resonate with the unique characteristics of each player type. This perspective not only allows for the development of more personalized narrative experiences but also offers game designers a way to experiment with storytelling that adapts to diverse player types. The tool can offer game designers new insights into how different players might engage with the game world and assist in redefining game narratives.

The paper is structured as follows. Section 2 reviews related work. Section 3 provides full technical details about the architecture, functionalities, and usage of ImageTeller. Section 4 reports the experiments conducted with ImageTeller. Section 5 presents concluding remarks.

## 2 Related Work

The concept of generating stories from visual input has been explored in some previous related works. Early research in this area, such as the work by Farhadi et al. [5], focused on generating descriptive sentences from images by comparing estimations of meaning derived from both visual and textual data. Subsequent studies explored the potential of AI models to create more elaborate stories based

on images or image sequences. For instance, Huang et al. [8] proposed a system that uses machine learning to generate descriptive captions for images and then constructs short stories from these captions. In a similar work, Smilevski et al. [28] introduced a Sequence-to-Sequence model with separate encoders for visual and narrative components, aiming to generate more human-like stories that go beyond basic image descriptions. Despite these advances, the generated stories tended to be limited in length and narrative richness, often failing to capture intricate plot structures or character arcs. The need for discourse coherence in visual storytelling is explored by Cardona-Rivera and Li [4], who introduced a system that generates stories around user-supplied photographs by considering discourse constraints during fabula generation.

Similarly, researchers have integrated player types into narrative generation to create personalized experiences. Thue et al. [30] introduced a player modeling approach for interactive storytelling, where narratives are dynamically adapted based on the player’s inferred style and preferences. By categorizing players according to their behavior, their system generates story content that aligns with individual player motivations. Orji et al. [23] investigated tailoring persuasive strategies in serious games to different player types, utilizing the BrainHex typology [21]. Their research demonstrated that customizing game content and narratives to match player types enhances player engagement and the effectiveness of persuasive messages. de Lima et al. [14] presented a method to interactive storytelling in games where quests and ongoing stories are determined based on individual personality and behavioral attitudes of the player, which are modeled using the Big Five personality traits [7].

While previous works have shown the benefits of using images for automated storytelling and integrating player types and personality traits into narrative generation, they often focus on short, descriptive narratives rather than longer, structurally complex stories. ImageTeller builds upon these previous efforts by not only generating richer and more engaging stories from images using the multi-modal capabilities of GPT-4o and its vision counterpart but also introducing the concept of player-driven narratives. By employing Bartle’s taxonomy of player types [1], ImageTeller generates stories that reflect diverse player motivations and behaviors, offering a novel approach to personalized narrative generation that combines visual input analysis with player type modeling.

### 3 The ImageTeller Prototype

ImageTeller was designed to explore the narrative potential of images by employing AI agents, namely GPT-4o, GPT-4o Vision, and FLUX.1 diffusion model, to analyze visual content and generate stories. Unlike purely text-driven approaches, ImageTeller’s integration of visual data aims to produce stories that harness both the imaginative potential of AI and the contextual richness of visual storytelling. The prototype is accessible at: <https://narrativelab.org/imageteller/>.

### 3.1 User Interface

The user interface of ImageTeller is presented in Fig. 1. To create a story, users can upload images on the system’s initial screen (Fig. 1(a)) by clicking on the “+” button (Action 1 in Fig. 1(a)). In the example illustrated in Fig. 1, the user has uploaded three screenshots from the game *The Legend of Zelda: Breath of the Wild* (Nintendo, 2017) without providing additional captions. After adding the images, the user can select a genre for the story. In this case, the user selected the “Explorer” player-driven genre (Action 2 in Fig. 1(a)).

Once the images are added and a genre is selected, the user can click on the “Generate Story” button to initiate the narrative generation process (Action 3 in Fig. 1(a)). The system then transitions to the story composition screen (Fig. 1(b)), where the user can visualize the generated story and its chapters, each accompanied by a title and an illustration.

For each chapter, the user has the option to either regenerate the illustration or regenerate the entire chapter by using the respective buttons available on the interface (Action 4 in Fig. 1(b)). These options allow the user to refine the narrative and its visual representation to better suit their preferences. At the bottom of the screen, the user can click on the “Save Story” button to save the generated story (Action 5 in Fig. 1(b)), which will then be added to the user’s personal library (Fig. 1(c)). The full story generated for the example presented in Fig. 1 is available at: <https://narrativelab.org/imageteller/#/story/394>.

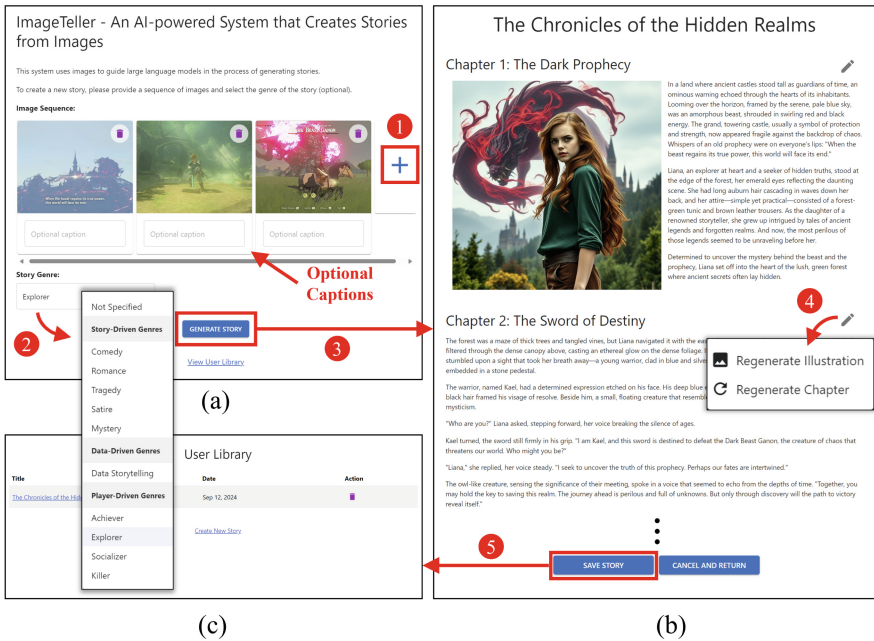
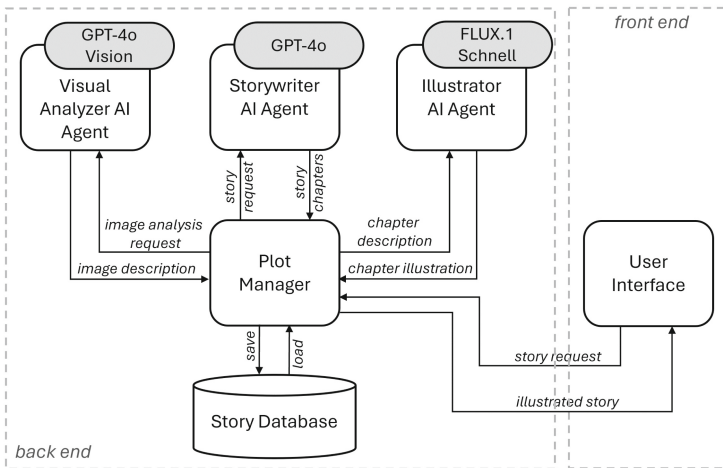


Fig. 1. The user interface of ImageTeller.

### 3.2 System Architecture

As illustrated in Fig. 2, the architecture of ImageTeller is based on the multi-AI-agent approach for narrative generation using large language models proposed in our previous works [11, 13, 18, 20, 26], which we adapted in the present work to support the generation of narratives from visual inputs. This architecture incorporates a set of AI agents, each with a distinct role in the narrative generation process: (1) a Visual Analyzer AI Agent, which uses the GPT-4o Vision to interpret and analyze the content of input images; (2) a Storywriter AI Agent, powered by the GPT-4o model, responsible for generating the textual narrative based on the results of the visual analysis; and (3) an Illustrator AI Agent that uses a text-to-image diffusion model to create visual representations that complement and enhance the story. A Plot Manager coordinates the interactions between agents and the narrative structure, similar to the plot management strategies used in interactive storytelling systems [16].

User interaction is facilitated through a user-friendly interface, allowing users to generate stories by providing images, captions, and selecting a genre. The interface displays the generated narratives in a format resembling illustrated books, with each chapter featuring a title, narrative text, and illustration. Users can provide interact by requesting alternative chapters or illustrations for iterative refinement of the results.



**Fig. 2.** The multi-AI-agent architecture of ImageTeller.

The AI agents are implemented using a plugin architectural approach. As indicated in Fig. 3, the Visual Analyzer AI Agent utilizes GPT-4o Vision via the OpenAI API for image analysis. The Storywriter AI Agent also utilizes GPT-4o model for text generation, while the Illustrator AI Agent is built upon a distilled version of the FLUX.1 text-to-image diffusion model (FLUX.1 Schnell), hosted

on a private server and accessed through a REST API. This modular design enhances the system’s adaptability, performance, and scalability.

### 3.3 Image Analysis

Analyzing and describing visual content is a core function of ImageTeller. This capability is powered by the advanced vision features of the GPT-4o model, which can extract detailed and contextually rich descriptions from each input image. These descriptions are produced using a prompt designed to extract and generate textual descriptions that are suitable for narrative generation.

The prompt used for image analysis (Prompt 1) is crafted to guide the model in identifying and describing key elements of the image, focusing on subjects, their actions, appearance, significant interactions, and environmental details that shape the scene’s mood and context. The description is generated as a single cohesive paragraph to ensure seamless integration of all relevant aspects.

**Prompt 1.** *Analyze the provided image and write a short description of the subjects present in the image, including their actions and appearance. Focus on key elements such as the interactions between subjects and any significant environmental details that contribute to the mood and context of the scene. Write the description as a single paragraph without using headers or lists.*

If a user opts to write a caption, it is directly incorporated into the analysis. The prompt is dynamically adjusted with the instruction: “*When generating the description, take into account that the image has the following caption:  $\langle C_i \rangle$ ”*”, where  $C_i$  represents the provided caption for image  $i$ . This allows the model to align the image analysis with the user’s desired interpretation, ensuring that the generated description is consistent with any specific narrative cues suggested by the caption. This feature is especially useful when users want to steer the narrative, specify character names, or highlight certain aspects of the image.

For example, when analyzing the second image used as input in the example illustrated in Fig. 1(a) (also shown in Fig. 3(b)), the model produced the description presented in Output 1, which captures the scene’s key elements and atmosphere.

**Output 1.** *The image depicts a character standing in a lush, mystical forest, with beams of sunlight filtering through the dense canopy above, casting a warm glow on the scene. The character, wearing a blue tunic and armor, is gripping the hilt of a legendary sword embedded in a stone pedestal, suggesting a moment of significant challenge or destiny. The surrounding environment features thick trees, verdant foliage, and the faint presence of mist, creating an atmosphere of ancient magic and mystery. The interplay of light, shadows, and greenery adds a sense of depth and tranquility, while the character’s determined stance and focused expression convey resolve and purpose.*

### 3.4 Story Generation

The story generation process in ImageTeller is designed to transform the visual and contextual data obtained from the image analysis into a cohesive and engaging narrative. This process is powered by the GPT-4o model, which is guided by a structured prompt system that varies depending on the user’s input, particularly regarding the type and genre of the narrative.

Three types of narratives are handled by ImageTeller: story-driven, player-driven, and data-driven. A *story-driven narrative* focuses on traditional storytelling and can either follow the conventions of a specific genre or be more general, without strict adherence to genre-specific elements. In our current prototype, we adopted the conventions of five fundamental genres that we have specified in a previous work [19]: Romance, Comedy, Tragedy, Satire, and Mystery. A *player-driven narrative* explores the motivations and preferences of different types of players according to Bartle’s taxonomy [1] (Achievers, Explorers, Socializers, and Killers). This type of story is designed to resonate with distinct player motivations by tailoring narrative elements—such as goals, challenges, exploration, social interactions, and conflicts—to align with the interests and behaviors of each player type. Table 1 presents the definitions of the player-driven genres according to Bartle’s taxonomy. Finally, a *data-driven narrative* is designed to emphasize the communication of insights and data points, prioritizing clarity and relevance while presenting complex information in an engaging way.

The prompt system used for the narrative generation process is composed of six modular components that are combined to create the final instruction set for the model. These components are:

1. **General Narrative Instruction** ( $P_{general}$ ), which offers core guidance for generating the story, including the structure, title format, and chapter division. This instruction ensures that the story follows a cohesive narrative flow and effectively brings the visual descriptions to life;
2. **Story Instruction** ( $P_{story}$ ), which focuses on transforming the image descriptions into a rich, character-driven narrative. This instruction guides the development of characters, their interactions, and the overall story dynamics, ensuring the narrative captures the essence of the scenes while maintaining a traditional storytelling style;
3. **Data Instruction** ( $P_{data}$ ), which shifts the focus of the narrative towards communicating insights and connections in a compelling way, tailoring the story to effectively present data-driven content with clarity and engagement;
4. **Story-Driven Genre Specification** ( $P_{genre}$ ), which defines the story genre and incorporates genre-specific elements such as setting, tone, and character types, ensuring the story adheres to the conventions of the chosen genre;
5. **Player-Driven Genre Specification** ( $P_{player}$ ), which adapts the narrative style and content to align with different player types according to Bartle’s taxonomy, ensuring the story reflects specific player motivations; and
6. **Image Descriptions** ( $P_{image}$ ), which contains the sequentially ordered descriptions of the images generated during the image analysis phase.

Table 2 presents the parameterized text for the six components of the prompt system. The final prompt used to guide the GPT-4o model in the narrative generation process is assembled by combining these components based on the user’s preferences:

1. For a story-driven narrative:  $P_{final} = P_{general} + P_{story} + P_{genre} + P_{image}$
2. For a player-driven narrative:  $P_{final} = P_{general} + P_{story} + P_{player} + P_{image}$
3. For a data-driven narrative:  $P_{final} = P_{general} + P_{data} + P_{image}$
4. If no genre is specified:  $P_{final} = P_{general} + P_{story} + P_{image}$

**Table 1.** Player-driven genre definitions based on Bartle’s taxonomy [1].

Player Type	Prompt Text
Achiever	<i>An Achiever player focuses on completing challenges, gaining rewards, and demonstrating mastery. Therefore, the narrative must be structured around goal-oriented tasks, measurable progress, milestones, and the desire for recognition. The protagonist is driven by accomplishment and overcoming increasingly difficult challenges</i>
Explorer	<i>An Explorer player is focused on discovery, curiosity, and unraveling hidden secrets. Therefore, the narrative must be structured around uncovering mysteries, exploring unknown locations, and the joy of finding something new. The protagonist is driven by a sense of wonder, investigation, and the thrill of the unknown</i>
Socializer	<i>A Socializer player focuses on relationships, interactions, and communication. Therefore, the narrative must be structured around friendships, alliances, community building, and the social dynamics between characters. The protagonist is driven by the need for connection, cooperative tasks, and shared experiences</i>
Killer	<i>A Killer player focuses on competition, dominance, and asserting power over others. Therefore, the narrative must be structured around conflict, strategic moves, and competition. The protagonist is driven by a desire for victory, control, and exerting influence over other characters</i>

A key feature of  $P_{general}$  is its directive that instructs the model to generate the narrative in the expected output format, with the story divided into chapters and markdown headers used to identify titles. This structured format ensures the narrative is organized with clear distinctions between the story title, chapter titles, and chapter content. By analyzing the markdown headers, the system can parse the generated narrative and process each chapter individually to produce the final formatted version of the narrative to be displayed to users. An example of a complete prompt and the raw narrative content generated by GPT-4o is presented in our complementary open-access paper [12].



**Table 2.** Components of the prompt system for narrative generation, where  $\langle g_i^{name} \rangle$  identifies the genre name or player type,  $\langle g_i^{description} \rangle$  represents the definition of the genre or player type, and  $\langle I_{description} \rangle$  is an ordered set of image descriptions.

Component	Prompt Text
$P_{general}$	<i>Using the following set of sequentially ordered image descriptions, write a cohesive, engaging, and complete story resembling a book narrative. The story must have a creative title written using markdown header level 1 (#). Follow the sequence of events as represented in the descriptions, maintaining consistency in mood and context. Ensure the narrative has a clear beginning, middle, and end. Divide the story into a natural number of chapters, each with a name and number, using markdown header level 2 (##) for the chapter titles</i>
$P_{story}$	<i>When writing the story, assign names to the subjects, describe their visual attributes (eye color, hair color, hairstyle) and clothes (type, color, style), and develop their characters based on their described actions and appearances. Ensure the story incorporates all key elements from the image descriptions, including the subjects, their interactions, and significant environmental details. Focus on creating a narrative and dialogue that brings the story to life, rather than merely describing the images. Exclude any details specific to the image format, such as speech balloons, text boxes, or floating text</i>
$P_{data}$	<i>Highlight key insights, connect the dots between different data points, and present a clear and engaging story. Use metaphors, analogies, and emotional appeal to make the story engaging. Avoid referring directly to the images or describing them. Instead, weave the data into the narrative naturally, as if explaining the insights to someone who has never seen the images. Also suggest potential implications, actions, or reflections based on the insights presented</i>
$P_{genre}$	<i>Incorporate genre-specific elements such as setting, tone, themes, and character types to clearly reflect the <math>\langle g_i^{name} \rangle</math> genre. Use the following definition for the <math>\langle g_i^{name} \rangle</math> genre: <math>\langle g_i^{description} \rangle</math></i>
$P_{player}$	<i>The protagonist’s journey presented in the narrative must be based on Bartle’s definition of the <math>\langle g_i^{name} \rangle</math> player type. Use the following definition for the <math>\langle g_i^{name} \rangle</math> player type: <math>\langle g_i^{description} \rangle</math></i>
$P_{image}$	<i>Image descriptions: <math>\langle I_{description} \rangle</math></i>

### 3.5 Image Generation

In ImageTeller, each story is structured into multiple chapters, with each chapter comprising a title, text, and an illustration. The process for generating illustrations for the chapters takes advantage of the recent advancements in text-to-image machine learning models, such as DALL-E, Midjourney, and Stable Diffusion. In our current implementation, we use the FLUX.1 Schnell model,<sup>1</sup> which

<sup>1</sup> <https://huggingface.co/black-forest-labs/FLUX.1-schnell>.

is a distilled version of the FLUX.1 text-to-image transformer model capable of generating high-quality images in 1 to 4 steps.

The generation of an illustration begins with the GPT-4o model identifying and describing a significant event in the chapter text. This description is then used as input for the image generation model. Prompt 2 shows the prompt used to instruct GPT-4o for this task, where  $C_i^{text}$  is the text of a chapter  $C_i$ .

**Prompt 2.** *Identify a single significant event in the provided story chapter and write a concise description (maximum of 80 words) to be used as input for a text-to-image model to generate an illustration. Focus on the characters' actions, interactions, and the environment. When describing the characters, include all details related to their visual attributes (you must describe their age group, hair color, hairstyle, eye color, clothing type, and clothing colors). Write only the description with no formatting elements. Story Chapter:  $\langle C_i^{text} \rangle$ .*

While directly providing the generated description to a text-to-image model would suffice for producing an illustrative representation for a chapter, we implemented additional prompt optimizations to improve image quality. First, we incorporated a default illustration style (“*Photography, realistic detailed skin, 4k, highly detailed, diffused soft lighting, shallow depth of field, sharp focus, hyperrealism, cinematic lighting*”), which is combined with the chapter description to define a consistent visual aesthetic for all generated images. Second, we added a general negative prompt to guide the model toward generating more accurate images by explicitly identifying undesirable visual elements in narrative events, as proposed in our previous work [20]. Third, we implemented a technique to maintain character consistency across consecutive images by replacing character names with those of well-known actors familiar to the text-to-image model. This approach ensures consistent character representations, enhancing the visual coherence of the chapter illustrations.

Prompt 3 shows the final prompt generated for the text-to-image model considering Chap. 1 of the story presented in Fig. 1(b) as input. The generated illustration is also shown in Fig. 1(b).

**Prompt 3.** *(Emma Watson:1.2) stands at the edge of a lush, green forest, her emerald green eyes fixed on the distant horizon. Her long auburn hair cascades in waves down her back, and she wears a forest-green tunic and brown leather trousers. Behind her, an amorphous beast, shrouded in swirling red and black energy, looms ominously over a towering castle, contrasting against the serene, pale blue sky. The atmosphere is tense, filled with a sense of impending danger. Photography, realistic detailed skin, 4k, highly detailed, diffused soft lighting, shallow depth of field, sharp focus, hyperrealism, cinematic lighting.*

## 4 A Fourfold Experiment on Player-Driven Narratives

To evaluate ImageTeller’s ability to generate narratives for each of Bartle’s player types [1], we conducted an experiment using the same sequence of three images

shown in Fig. 3 as input. These images are screenshots captured from *The Legend of Zelda: Breath of the Wild*, a Nintendo video game known for its rich storytelling [29]. For each player type, we generated a narrative and analyzed its adherence to narratology principles and alignment with the characteristics of that type, drawing parallels with well-known video game protagonists who exemplify each player type.



**Fig. 3.** The screenshots from *The Legend of Zelda: Breath of the Wild* used in the experiment: (a) Calamity Ganon circling around Hyrule Castle; (b) Link drawing the Master Sword from its pedestal; and (c) Link fighting against Dark Beast Ganon.

#### 4.1 Achiever-Oriented Story - The Quest for Glory

- Resulting story: <https://narrativelab.org/imageteller/#!/story/395>.

The generated story is a straightforward instance of Frye’s romance genre [6], with an epic plot [25], narrating the quest of Cedric, a protagonist charismatically marked by his unique ability to extract and wield the legendary Sword of Light—aligning with Propp’s 17th function [24] and reminiscent of King Arthur’s Excalibur episode [10]. Driven by the pursuit of achievement, Cedric faces the challenge of preventing the rise of the Beast of Shadows. His victory not only saves his world but also earns him the recognition he always sought.

The narrative clearly aligns with the achiever player type, focusing on accomplishment, mastery, and tangible rewards. As Lacy discusses [9], the term “achieving”, in particular in the expression “achievement of the Grail”, had a double sense in the original French romances, since the French word “achever” conveyed both the success in accomplishing the goal of the Grail quest, and the termination of the related adventures. Similarly, the generated story concludes with Cedric’s personal triumph and the realization that no challenges remain, satisfying the Achiever’s desire for both accomplishment and closure. This resembles the original journey of Link in *The Legend of Zelda* series, where Link undertakes epic quests to defeat evil forces, mastering skills and acquiring powerful items to save Hyrule, ultimately gaining recognition as a legendary hero.

#### 4.2 Explorer-Oriented Story - The Chronicles of the Hidden Realms

- Resulting story: <https://narrativelab.org/imageteller/#!/story/394>.

The generated narrative introduces a female protagonist, Liana, the daughter of a renowned storyteller, leading to an incursion into epistemic plots involving mystery narratives [25]. Determined to uncover the mystery behind the calamitous beast and the prophecy, she embodies the explorer player type by focusing on discovery and unraveling hidden truths. Instead of the fragile victim stance typical of the heroic monomyth pattern [3,31], Liana combines the auxiliary hetaira role—guiding the sword-bearer with her exploratory talents—and the active amazon role—aggressively participating in the successful combat—as in the feminine psyche proposal of one of Carl Jung’s collaborators [32].

The narrative aligns with the explorer player type, emphasizing curiosity, exploration, and the thrill of uncovering secrets. The story satisfies the explorer’s desire for new experiences and knowledge, mirroring modern heroic feminine renditions such as Lara Croft in the *Tomb Raider* game series, who embodies the spirit of exploration by venturing into unknown territories, uncovering ancient artifacts, and solving mysteries driven by her curiosity and determination.

### 4.3 Socializer-Oriented Story - The Bonds of Destiny

- Resulting story: <https://narrativelab.org/imageteller/#/story/404>.

The generated story features a female protagonist, Elara, introduced as a “weaver of connections”, suggesting she ranks high in extraversion and agreeableness according to personality trait theories [7]. To fulfill her calling she ought to be particularly skilled in acts of verbal communication, the type of action that characterizes dramatic plots [25]. By allying herself with the sword-bearer, she gently persuades him to recruit a team of helpers, reflecting Propp’s 14th function (F9: “various characters place themselves at the disposal of the hero”) [24] and Vogler’s sixth stage (“Tests, Allies, Enemies”) [31]. The inclusion of an old sage, corresponding to Vogler’s fourth stage (“Meeting with the Mentor”) [31], emphasizes themes of trust and unity, which the protagonist rephrases into an appeal that each one of them should be “the light for each other”. As the beast is subdued, Elara’s greatest reward is seeing her friends “beaming with relief and triumph”, highlighting the importance of social bonds.

The plot directly aligns with the socializer player type by emphasizing relationships, cooperation, and communal success. It satisfies the Socializer’s desire for connection and shared experiences, demonstrating how unity and trust lead to overcoming challenges. The story integrates social interaction as the core mechanism for achieving victory, consistent with Bartle’s definition of socializer motivations [1]. This focus on building bonds and working together is a core narrative element in the *Mass Effect* game series, where the protagonist, Commander Shepard, forms alliances with a diverse crew, and their collective efforts and relationships are crucial to overcoming the galaxy’s threats.

### 4.4 Killer-Oriented Story - The Conqueror’s Path

- Resulting story: <https://narrativelab.org/imageteller/#/story/406>.

The male protagonist in the generated narrative, Roran, is singularly driven by a competitive desire to dominate, embodying Roger Caillois’s concept of the “agon” form of play [2], where a player obsessively seeks to outperform others in qualities like strength, skill, and ingenuity without external assistance, so as to defeat all adversaries. Described as having spent years “honing his skills, driven by an insatiable hunger to dominate and manipulate forces beyond human ken,” he exhibits no inclination toward socialization or collaboration. This mirrors the legendary figure of Roland, who refused to summon aid during the Battle of Roncevaux, relying solely on his prowess and his famed Durendal sword [27].

The story directly aligns with the killer player type, emphasizing competition, dominance, and personal power. After his solitary victory, the protagonist feels that “the path of the conqueror was his, and he would let nothing stand in his way”, illustrating the Killer’s desire for control and unchallenged supremacy. A contemporary parallel can be drawn with the protagonist of the *Doom* game series. Like Roran, Doom Slayer is a lone warrior driven by personal motivations to wage a one-man war against demonic forces.

## 5 Concluding Remarks

In this paper, we explored the generation of player-driven narratives from visual inputs. By employing AI agents that integrate advanced language models with visual analysis capabilities, we showed how ImageTeller can produce cohesive and engaging stories tailored to different player types, aligning with Bartle’s taxonomy [1]. The generated stories can serve as a source of inspiration, offering game designers new insights into how different players might engage with the game world and providing a helpful resource for refining game narratives.

Our initial experiments with ImageTeller, detailed in our complementary open-access paper [12], suggest that analyzing image sequences using the prototype provides a practical and effective approach to generating coherent narratives. The fourfold experiment presented in this paper demonstrated that ImageTeller effectively generates distinct narratives that resonate with the specific characteristics of the targeted player type. The analysis revealed that the generated stories not only adhere to established narratology principles but also resemble well-known narratives, including game franchises, reinforcing the relevance and potential applications of this method in the context of game design.

While the results are promising, there are limitations to consider. Our analysis of the results is primarily technical, focusing on alignment with narratology principles and player types, but comprehensive user studies are needed to assess the approach’s effectiveness in practical settings. A more in-depth investigation into the practical applications and usability of this method for assisting game designers will allow us to fully understand its potential benefits and challenges within the game industry. Additionally, the current implementation focuses on a predefined set of player types and genres. Expanding the tool to encompass a broader spectrum of player types taxonomies, such as the BrainHex typology [21], could enhance its versatility and applicability in game design scenarios.

As future work, we intend to conduct extensive user studies to refine the prototype. Such studies will help us better understand user preferences, improve the quality of the generated content, and tailor the system's outputs to meet the needs of both players and designers. Additionally, integrating user modeling techniques presents a promising avenue for future research. By guiding the story generation process according to user preferences for narrative content [15, 17] or personality traits [14], we expect to enhance personalization and engagement.

## References

1. Bartle, R.: Hearts, clubs, diamonds, spades: players who suit muds. *J. MUD Res.* **1**(1) (1996)
2. Caillois, R.: *Man, Play, and Games*. University of Illinois Press, Champaign (2001)
3. Campbell, J.: *The hero with a thousand faces*. New World Library (2008)
4. Cardona-Rivera, R., Li, B.: Plotshot: generating discourse-constrained stories around photos. In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 12, pp. 2–8 (2021). <https://doi.org/10.1609/aiide.v12i1.12860>
5. Farhadi, A., et al.: Every picture tells a story: generating sentences from images. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010*. LNCS, vol. 6314, pp. 15–29. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15561-1\\_2](https://doi.org/10.1007/978-3-642-15561-1_2)
6. Frye, N.: *Anatomy of Criticism*. Princeton University Press, Princeton (2020)
7. Goldberg, L.R.: An alternative “description of personality”: the big-five factor structure. *J. Pers. Soc. Psychol.* **59**(6), 1216–1229 (1990). <https://doi.org/10.1037/0022-3514.59.6.1216>
8. Huang, T.H.K., et al.: Visual storytelling. In: Knight, K., Nenkova, A., Rambow, O. (eds.) *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1233–1239. Association for Computational Linguistics, San Diego, California (2016). <https://doi.org/10.18653/v1/N16-1147>
9. Lacy, N.J.: *The Grail, the Quest, and the World of Arthur*. Boydell & Brewer, Martlesham (2008)
10. Lacy, N.J., Wilhelm, J.: *The Romance of Arthur*. Routledge, Milton Park (2012)
11. de Lima, E.S., Casanova, M.A., Feijó, B., Furtado, A.L.: Semiotic structuring in movie narrative generation. In: Ciancarini, P., Iorio, A.D., Hlavacs, H., Poggi, F. (eds.) *Entertainment Computing – ICEC 2023*, pp. 161–175. Springer, Singapore (2023). [https://doi.org/10.1007/978-981-99-8248-6\\_13](https://doi.org/10.1007/978-981-99-8248-6_13)
12. de Lima, E.S., Casanova, M.A., Furtado, A.L.: Imagining from images with an AI storytelling tool. In: *ArXiv e-prints*, arXiv: 2408.11517 [cs.CL] (2024). <https://doi.org/10.48550/arXiv.2408.11517>
13. de Lima, E.S., Feijó, B., Casanova, M.A., Furtado, A.L.: Chatgeppetto - an AI-powered storyteller. In: *Proceedings of the 22nd Brazilian Symposium on Games and Digital Entertainment*, pp. 28–37. ACM (2024). <https://doi.org/10.1145/3631085.3631302>
14. de Lima, E.S., Feijó, B., Furtado, A.L.: Player behavior and personality modeling for interactive storytelling in games. *Entertainment Comput.* **28**, 32–48 (2018). <https://doi.org/10.1016/j.entcom.2018.08.003>

15. de Lima, E.S., Feijó, B., Furtado, A.L.: Adaptive storytelling based on personality and preference modeling. *Entertainment Comput.* **34**, 100342 (2020). <https://doi.org/10.1016/j.entcom.2020.100342>
16. de Lima, E.S., Feijó, B., Furtado, A.L.: Managing the plot structure of character-based interactive narratives in games. *Entertainment Comput.* **47**, 100590 (2023). <https://doi.org/10.1016/j.entcom.2023.100590>
17. de Lima, E.S., Feijó, B., Furtado, A.L., Gottin, V.M.: Personality and preference modeling for adaptive storytelling. In: 2018 17th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), pp. 187–18709 (2018). <https://doi.org/10.1109/SBGAMES.2018.00030>
18. de Lima, E.S., Neggers, M.M.E., Casanova, M.A., Feijó, B., Furtado, A.L.: A pattern-oriented AI-powered approach to story composition. In: Figueroa, P., et al. (eds.) ICEC 2024, pp. 1–16. Springer, Cham (2024). [https://doi.org/10.1007/978-3-031-74353-5\\_11](https://doi.org/10.1007/978-3-031-74353-5_11)
19. de Lima, E.S., Neggers, M.M.E., Furtado, A.L.: Multigenre AI-powered story composition. In: ArXiv e-prints, [arXiv: 2405.06685](https://arxiv.org/abs/2405.06685) [cs.CL] (2024). <https://doi.org/10.48550/arXiv.2405.06685>
20. de Lima, E.S., Neggers, M.M., Feijó, B., Casanova, M.A., Furtado, A.L.: An AI-powered approach to the semiotic reconstruction of narratives. *Entertainment Comput.* **52**, 100810 (2025). <https://doi.org/10.1016/j.entcom.2024.100810>
21. Nacke, L.E., Bateman, C., Mandryk, R.L.: Brainhex: a neurobiological gamer typology survey. *Entertainment Comput.* **5**(1), 55–62 (2014)
22. Oktaviana, A.A., Joannes-Boyau, R., Hakim, B., Burhan, B., Sardi, R., Adhityatama, S., et al.: Narrative cave art in Indonesia by 51,200 years ago. *Nature* **631**(8022), 814–818 (2024). <https://doi.org/10.1038/s41586-024-07541-7>
23. Orji, R., Vassileva, J., Mandryk, R.L.: Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Model. User-Adap. Inter.* **24**(5), 453–498 (2014)
24. Propp, V.: *Morphology of the Folktale*. University of Texas Press, Austin (1968)
25. Ryan, M.-L.: Interactive narrative, plot types, and interpersonal relations. In: Spierling, U., Szilas, N. (eds.) ICIDS 2008. LNCS, vol. 5334, pp. 6–13. Springer, Heidelberg (2008). [https://doi.org/10.1007/978-3-540-89454-4\\_2](https://doi.org/10.1007/978-3-540-89454-4_2)
26. Schetinger, V., Bartolomeo, S.D., de Lima, E.S., Meinecke, C., Rosa, R.: *n* walks in the fictional woods. In: ArXiv e-prints, [arXiv: 2308.06266](https://arxiv.org/abs/2308.06266) [cs.HC] (2023). <https://doi.org/10.48550/arXiv.2308.06266>
27. Scott-Moncrieff, C.K.: *The Song of Roland*. The Project Gutenberg (1996). <https://gutenberg.org/cache/epub/391/pg391-images.html>
28. Smilevski, M., Lalkovski, I., Madjarov, G.: Stories for images-in-sequence by using visual and narrative components. In: Kalajdziski, S., Ackovska, N. (eds.) ICT 2018. CCIS, vol. 940, pp. 148–159. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00825-3\\_13](https://doi.org/10.1007/978-3-030-00825-3_13)
29. Stark, C.: Breath of the wild is one of the best games of this decade, and perhaps the most impactful of the next decade (2019). <https://www.polygon.com/2019/11/11/20955542/legend-of-zelda-breath-of-the-wild-best-games-decade>
30. Thue, D., Bulitko, V., Spetch, M., Wasylshen, E.: Interactive storytelling: a player modelling approach. In: Third Artificial Intelligence and Interactive Digital Entertainment Conference, pp. 43–48 (2007). <https://doi.org/10.1609/aiide.v3i1.18780>
31. Vogler, C.: *The Writer's Journey*. Michael Wiese Productions (2007)
32. Wolff, T.: *Structural Forms of the Feminine Psyche*. C. G. Jung Institute (1956)