Generative-AI-based game asset creation: developing a system for supporting game production brainstorming

Luiz Paulo S. B. Cavalcanti[1], **Pamela T. L. Bezerra**[1], **Gabriella A. B. Barros**[2]

[1]Faculdade do Centro de Estudos e Sistemas Avançados do Recife – CESAR School
Avenida Cais do Apolo, n° 77 – 50.030-220 – Recife – PE – Brasil

{lpsbc,ptlb}@cesar.school

[2]Independent Professsional
Alagoas, Brasil

gabbbarros@gmail.com

Abstract. Introduction*: Recent advances in generative AI have surprised the public by quickly generating text and images of good quality, impacting many entertainment industries, like the game industry. Despite concerns related to data privacy and plagiarism, generative AI can support artists and developers during the brainstorming process, enhancing creativity and performance.* ***Objective:*** *This research develops a system that integrates different generative AI technologies to assist game asset creation (text, image, and sound) all in one place. Developers can define, among other things, a protagonist's concept art, generate associated NPCs (both image and text), create quests (text), and a short soundtrack (audio).* ***Methodology or Steps****: To develop the proposed solution, an initial study on the different generative AI models for different assets was conducted. This was followed by the system development using Python, Streamlit, and some APIs. Finally, a set of experiments was conducted to analyze the system's capability to generate high-quality and diverse assets for a role-playing game.* ***Results:*** *The results demonstrate the system's potential to facilitate the creative process by optimizing the prompt engineering process and combining multiple generative AI technologies in a single tool, offering a versatile solution for both small teams and larger studios.*

Keywords *Generative AI models, Assets generation, Game Development, Game Art Generation.*

1. Introduction

Game development can entail substantial financial costs, particularly for smaller studios seeking to produce assets such as realistic 3D art. These monetary constraints hinder their ability to compete with larger studios and publishers. For example, independent studios, typically composed of teams ranging from five to ten members, encounter significant challenges, with development costs estimated between USD 50,000 and USD 750,000 [Auroch Digital 2022]. Additionally, the whole pipeline can rely on a variety of assets, including 2D or 3D art, text, and audio. These assets pass through multiple stages, from conceptualization to final implementation, and not all assets initially created make it to the final product [GDC 2017].

Meanwhile, recent advances in artificial intelligence (AI) have led to the development of generative models. These models can generate new data based on

patterns learned from a dataset, a great improvement compared to discriminative models, which can only categorize data [Géron 2022]. Large Language Models (LLMs), such as OpenAI's chatGPT [Kalyan 2024], are examples of generative models capable of producing human-like text.

Economy and industry reports, such as Deloitte's 2024 *Gaming Outlook Report* [1], discuss how generative AI is being leveraged to address rising development costs and drive innovation. While further empirical evidence is needed, early applications of generative AI in game production include concept art iteration [2], coding assistance, conversational Non-Playable Characters (NPCs) [Proxima Enterprises Inc 2024], translation [Future Trans 2024], game levels, and the creation of digital assets [Schrum et al. 2020]. As technology evolves, defining a coherent implementation strategy remains a challenge, but existing research and technical reports continue to document its expanding role in different phases of game development.

Considering this context, this paper proposes a **system for integrating multiple generative models to support game asset creation** during the initial game development stages, especially brainstorming. The proposed solution's main goals are (1) to consolidate the generation of various assets, such as text, image, and sound, in one system; and (2) to reduce the prompt engineering process, enabling, therefore, designers and developers to focus on the creative process instead of focusing on the technicalities of each model. Additionally, by combining LLMs with image and sound generation models, a wide variety of characters' concept arts, storylines, and soundtracks, for example, can be explored and combined, expanding the creative possibilities of game development.

This paper is organized as follows: Section 2 discusses solutions of interest; Section 3 details the system developed; Section 4 describes the the system's evaluation and results; and finally Section 5 discusses current results and following research steps.

2. Related Works

Generative AI tools have advanced in the text, image, and audio domains, offering new capabilities for game development. In terms of **text generation**, LLM models like GPT-4 [OpenAI 2023], LLaMa [Meta AI 2023b], Gemini [DeepMind 2024b], and Deep Seek [DeepMind 2024a] generate context-aware language for diverse applications. For **image creation**, Stable Diffusion [Esser et al. 2024] supports fine-tuned generation, while Layer AI [AI 2023] integrates with Unity to streamline 2D and audio asset creation. Finally, for **audio production**, tools such as Audiocraft/MusicGen [Meta AI 2023a] and AIVA [Technologies 2022] generate music from text prompts, combining AI and procedural methods.

These tools have been increasingly adopted in game development to boost productivity and reduce costs [Deloitte 2024]. They support a wide range of tasks, from concept art and illustrations [Earle et al. 2024] to narrative design, including stories and quests [Mao et al. 2024]. This versatility streamlines artistic workflows and enables autonomous content creation, fostering creativity and allowing small teams to develop prototypes [JaeJun Lee 2023].

[1] https://www2.deloitte.com/content/dam/Deloitte/us/Documents/technology-media-telecommunications/us-2024-me-outlook-generative-ai.pdf

[2] https://gameanalytics.com/blog/creating-concept-art-for-games-with-ai/

Meanwhile, **Generative AI research** has expanded these models' role in game development. For example, from art to narrative design, LLMs have been studied and applied to dynamic storytelling, NPC interaction [Buongiorno et al. 2024], and procedural quest generation [Ammanabrolu et al. 2020]. On the other hand, Transformer-based systems enable adaptive music generation aligned with gameplay [dos Santos et al. 2022].

Prompt engineering techniques, another important topic of study, further enhance AI's utility in game design. Projects like *Worldsmith* [Dang et al. 2023] introduce iterative prompting methodologies to facilitate immersive world-building, while Sun et al. present a co-creative storytelling game that leverages generative AI to collaboratively construct interactive narratives [Sun et al. 2023].

Nevertheless, despite these advances, few existing works offer integrated frameworks capable of simultaneously generating 2D art, text, and audio for games, which is the purpose of the research.

3. Proposed Solution

This work presents a **multi-modal content generation system** tailored for the ideation and prototyping of game character assets. The system LLMs, image generation, and simple audio synthesis tools into a single web interface. It allows users to interactively generate and manage characters, NPCs, quests, and narrative content for world-building applications.

The tool enables four key functionalities:

1. **Character**: Generate a character profile (name, gender, race, class, background) and backstory using Phi 3.5 Mini.
2. **Images**: Create character portraits with different styles (e.g., pixel-art, anime) using DALL-E.
3. **NPCs and Quests**: Produce NPCs and quest descriptions based on the character using Phi 3.5 Mini.
4. **Audio**: Music theme based on the character's description using Audiocraft.

The system was implemented in Python in Streamlit Labs [3]. **Figure 1** shows a portion of the interface, highlighting the character creation process. The following section describes each step in this process.

3.1. System Workflow

Figure 2 details the system workflow, starting with user input of a character's **name** and **gender**, while traits like **race**, **class**, and **background** can be selected or randomly assigned from predefined fantasy archetypes. An LLM (GPT-4o-mini) generates a backstory and origin text, which guides the creation of visual assets.

Next, the system produces up to **10 images** in various styles (e.g., anime, 8-bit) using DALL·E, including full-body portraits, turn-around views, and scenes of the character's origin. Users can further expand the narrative by generating **supporting NPCs** with backstories, alongside **quests**—short narrative tasks involving the main character.

[3]Code available at: `https://tinyurl.com/yavwupsh`

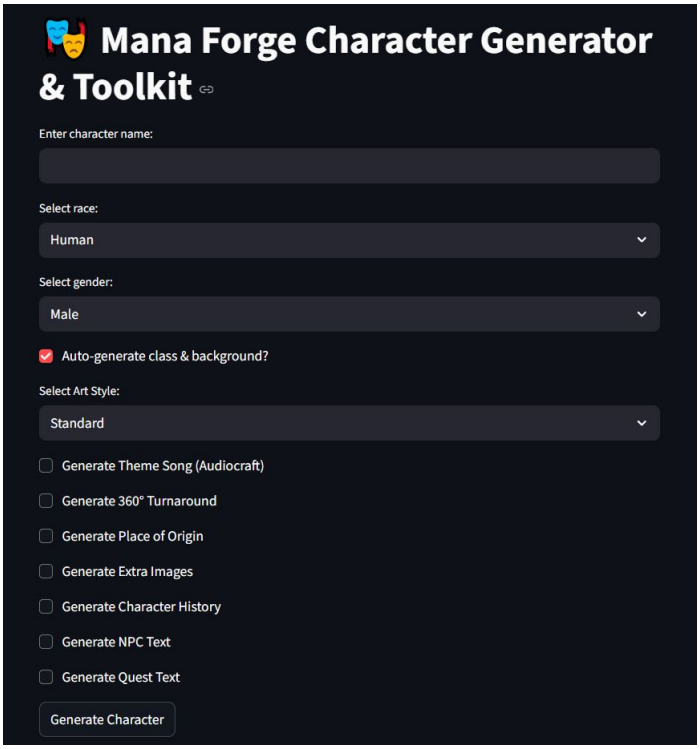

Figura 1. System's web interface, showing character creation options.

Finally, the character's description is used to generate different soundtracks, each one with a melody that fits the character's description.

In summary, the system generates the following outputs:

- Up to 10 stylized character images (full-body, 360°, origin)
- Supporting NPCs and quest narratives
- A thematic soundtrack created from the character's description
- All these assets can be exported in JSON or PDF format.

4. Experiments

A series of experiments evaluated the system's ability to generate coherent assets within a unified fantasy theme, integrated into a functional prototype of a medieval fantasy card game developed in Unity. The prototype features 48 characters (e.g., warriors, mages, and rogues from different races), each with images, narrative text, and individual soundtracks, resulting in **718 images, 48 PDFs, and 78 audio tracks**. Assets were assessed based on visual fidelity, narrative coherence, and thematic alignment of audio. All generated content is available at the link at the bottom of this page[4].

4.1. Summary of Results

Image results showed visually coherent portraits aligned with race, class, and profession. Based on visual inspection, around **86%** of the 718 images were considered suitable for in-game use. Minor issues, such as proportion or perspective errors, were observed in about **7%**.

[4]https://tinyurl.com/4ydudanp

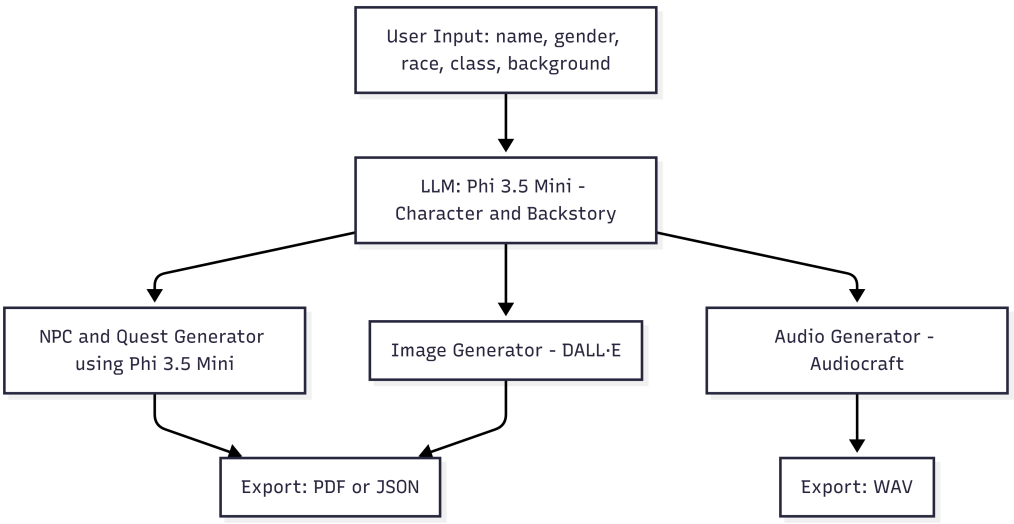

Figura 2. System's workflow describing the assets generated and models used.

Text results showed that approximately **82%** of the generated NPCs and quests maintained narrative coherence and alignment with character traits. Minor encoding or consistency issues were observed in around **30%** of the outputs,

Sound results were less consistent: only **61%** of the 78 tracks were considered thematically appropriate and free of noticeable artifacts. About **39%** had technical issues (e.g., glitches, abrupt cuts), and **24%** lacked musical coherence or complexity.

Overall, the system produced coherent and thematically aligned assets across modalities, with audio generation identified as the area requiring the most improvement.

5. Conclusions and Future Work

This paper presents a system integrating multiple generative models to create diverse game assets, aiming to support game development teams by automating prompt engineering and fostering human-AI collaboration in brainstorming and prototyping stages. Experiments demonstrated its potential to generate coherent and thematically consistent text, images, and audio, suitable mainly for early development phases. Limitations include processing time—especially for audio—and asset quality below final production standards.

Future work will focus on improving audio generation by extending soundtrack complexity and exploring alternative text and image models to enhance asset quality. Developing a user-friendly web interface will support better input and visualization. Ethical challenges related to originality, ownership, bias, and impact on creative professions will also be investigated. Finally, user testing with artists and developers will evaluate the system's usability and effectiveness across diverse backgrounds.

Referências

AI, L. (2023). Layer ai: Customizable image generation for 3d assets. `https://www.layer.ai/`.

Ammanabrolu, P., Broniec, W., Mueller, A., Paul, J., e Riedl, M. O. (2020). Toward automated quest generation in text-adventure games.

Auroch Digital, I. (2022). How much does it cost to make a game. https://www.aurochdigital.com/blog/2021/8/19/how-much-does-it-cost-to-make-a-game [Accessado em: (12/01/2025)].

Buongiorno, S., Klinkert, L., Zhuang, Z., Chawla, T., e Clark, C. (2024). Pangea: procedural artificial narrative using generative ai for turn-based, role-playing video games. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 20, pages 156–166.

Dang, H., Brudy, F., Fitzmaurice, G., e Anderson, F. (2023). Worldsmith: Iterative and expressive prompting for world building with a generative ai. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–17.

DeepMind (2024a). Deep seek: Advanced search and synthesis for textual information.

DeepMind, G. (2024b). Gemini: A multimodal approach for advanced ai tasks.

Deloitte (2024). Deloitte 2024 gaming outlook: Transmedia and generative ai. Accessed: 2025-01-26.

dos Santos, G. A. C., Baffa, A., Briot, J.-P., Feijó, B., e Furtado, A. L. (2022). An adaptive music generation architecture for games based on the deep learning transformer model. *arXiv preprint arXiv:2207.01698*.

Earle, S., Parajuli, S., e Banburski-Fahey, A. (2024). Dreamgarden: A designer assistant for growing games from a single prompt.

Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., Levi, Y., Lorenz, D., Sauer, A., Boesel, F., et al. (2024). Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first international conference on machine learning*.

Future Trans (2024). Leveling up: How generative ai is revolutionizing game localization. Available at https://future-trans.com/how-generative-ai-is-revolutionizing-game-localization/. Accessed on January 28, 2025.

GDC, I. (2017). Achieving two worlds, every year: How magic the gathering sustainably doubled its worldbuilding. https://www.youtube.com/watch?v=9LijlGLFKTE [Accessado em: (02/02/2025)].

Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. "O'Reilly Media, Inc.".

JaeJun Lee, So-Youn Eom, J. L. (2023). Empowering game designers with generative ai. *International Journal of Computer Science and Information Systems*.

Kalyan, K. S. (2024). A survey of gpt-3 family large language models including chatgpt and gpt-4. *Natural Language Processing Journal*, 6:100048.

Mao, X., Yu, W., Yamada, K. D., e Zielewski, M. R. (2024). Procedural content generation via generative artificial intelligence.

Meta AI (2023a). Audiocraft: Ai music generation. https://ai.facebook.com/blog/audiocraft-ai-generated-music/.

Meta AI (2023b). Llama: Open and efficient foundation language models. https://arxiv.org/abs/2302.13971.

OpenAI (2023). Gpt-4. https://openai.com/research/gpt-4.

Proxima Enterprises Inc (2024). Suck up! Digital. Available at: https://www.playsuckup.com/. Accessed on January 28, 2025.

Schrum, J., Gutierrez, J., Volz, V., Liu, J., Lucas, S., e Risi, S. (2020). Interactive evolution and exploration within latent level-design space of generative adversarial networks.

Sun, Y., Li, Z., Fang, K., Lee, C. H., e Asadipour, A. (2023). Language as reality: a co-creative storytelling game experience in 1001 nights using generative ai. 19(1):425–434.

Technologies, A. (2022). Aiva: Artificial intelligence virtual artist for music composition. https://www.aiva.ai/.