CrawLLM: Theming Games with Large Language Models

Marvin Zammit
University of Malta
Msida, Malta
marvin.zammit@um.edu.mt

Antonios Liapis
University of Malta
Msida, Malta
antonios.liapis@um.edu.mt

Georgios N. Yannakakis
University of Malta
Msida, Malta
georgios.yannakakis@um.edu.mt

***Abstract*—Video game players increasingly seek immersive and personalised experiences that resonate with their unique interests and personalities. This study explores the novel application of large language models (LLMs) in game re-theming, a process that adapts game assets to new settings and narratives. We applied this to an original game, CrawLLM, which combines dungeon crawling and card combat mechanics. The gameplay structure guided the construction of prompts for LLMs to generate new themes, stories, characters, and locations, which then directed the generation of corresponding visual assets. This approach enabled the game's aesthetics to emerge from its underlying narrative. We hereby demonstrate a playable version of the game with 20 pre-generated, non-curated themes. This study paves the way for future research on automated game generation, where LLMs can orchestrate diverse content creation pipelines to construct entire games from high-level prompts, while maintaining cohesive user experiences.**

***Index Terms*—games, procedural content generation, image generation, text generation, large language models**

I. Introduction

In the ever-evolving landscape of video games, developers are frequently faced with the challenge of creating highly immersive gaming experiences that resonate with a wide range of players with diverse interests and personalities. Traditional game development processes often adhere to predefined visual assets and storylines, limiting the capacity for dynamic adaptation and player-driven customisation. Content generators for individual game facets are abundant, but the orchestration of multiple facets required for a cohesive outcome is a difficult and as yet unsolved problem [1], despite some notable work in the field [2]. The recent advances in large language models (LLMs) have endowed them with the ability to comprehend and interpret complex game concepts, narratives, and themes [3]. Their extensive training on vast textual corpora equips them with a rich understanding of language, enabling them to grasp the nuances and intricacies necessary to generate adaptations, whilst maintaining thematic coherence. Combined with the capabilities of text-to-image generative models, such as Stable Diffusion (SD) [4], LLMs can orchestrate a sophisticated content generation pipeline, enabling the dynamic re-theming of games by changing their narrative and visual assets. We hereby explore how LLMs may be used as a top-down pipeline for orchestration [1] to adapt and personalise gaming experiences.

II. CrawLLM

To test the versatility of our proposed system, we developed CrawLLM, a dungeon crawler game using the Unity game engine[1]. The player navigates top-down mazes, battling enemies and collecting keys to unlock doors. The player's goals is to find the exit. Doors require varying numbers of keys, may be one-way, or hidden. Combat is card-based, where players select action cards using mana. Winning a combat round grants new cards for future encounters. Due to the time-intensive nature of sequentially generating multiple assets, we present a playable demonstration featuring 20 pre-generated themes constituting a complete generative run, rather than a curated selection. The dungeon layout and solvable door/key puzzles are dynamically generated using grammars [5] built around the unpublished cyclic dungeon generation method from *Unexplored* (Ludomotion, 2017). The dungeon is divided into cycles of rooms, each representing a narrative location with its own door/key puzzle. Rooms may be replaced by sub-cycles, to increase complexity while retaining solvability.

III. Theme Generation

The Mixtral 8x7B LLM [6] was used to generate new themes. Figure 1 shows the iterative sequence of narrative elements, names, and descriptions which were generated. In addition to the theme, main plot, and protagonist name and descriptions, the names of locations for the dungeon cycles are generated. In each location, a main antagonist and subsequently two minions are generated to embellish the combat encounters and add difficulty. Furthermore the LLM produces the prompts required to create visual assets via a Stable Diffusion XL (SDXL) [7] text-to-image generative model. Figure 2 shows examples of the game's visuals with different themes applied.

The game structure posed several challenges for content generation. Tile set images needed visual coherence yet distinction across dungeon locations. Creating subtle secret door

This project has received funding from the European Union's Horizon 2020 programme under grant agreement No 951911.

979-8-3503-5067-8/24/$31.00 ©2024 IEEE

[1] https://unity.com/

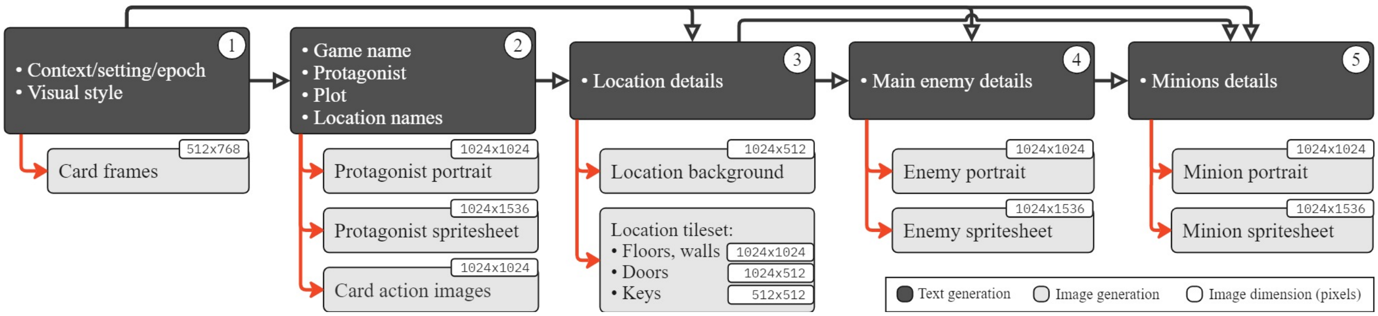
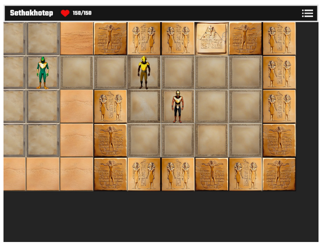
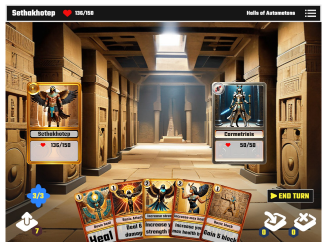


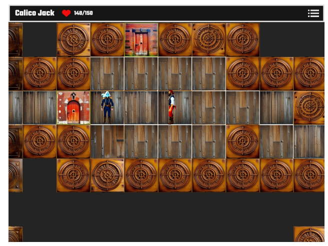
Fig. 1: The text and image assets generation sequence. Pixel dimensions for images are also shown. Arrows delineate the inputs used to generate the respective assets. Outputs from block #1 were also used in all image prompts (arrows not shown).



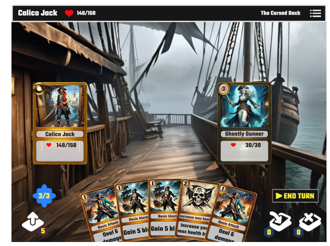
(a) Theme #1 dungeon



(b) Theme #1 card combat



(c) Theme #2 dungeon



(d) Theme #2 card combat

Fig. 2: Screenshots showing the dungeon crawling (a and c) and card combat sections (b and d) of CrawLLM with an ancient Egypt (Theme #1) and a pirate theme (Theme #2).

tiles from regular walls was difficult. This was accomplished by supplementing SDXL with ControlNET [8] models, providing rough outline sketches alongside text prompts to guide image generation. Card frames were similarly constrained by demarcating areas for text and images. To animate the characters' movement, ControlNET was provided with manually generated templates of poses and depth maps for each required frame. This ensured a precise positioning of the characters within the spritesheets, which were generated as one image for consistency. Card images, player portraits, and background images used a plain SDXL text-to-image pipeline.

IV. Discussion

The pipeline presented in this demo features no manual post-processing or curation of the generated assets, highlighting that these are fit for use despite some imperfections. However there are a number of issues with the current process. Firstly, the pipeline and the LLM prompts have to be meticulously crafted and tuned to match the game mechanics and individual assets required. Character consistency between different SD generations is also challenging, so multiple images of the same characters often result in notable visual differences. Furthermore enemies were limited to a humanoid form since bipedal pose templates were used to generate spritesheets. Another limitation is that pretrained open source SD models lack a transparency channel, requiring automated background removal which often omits peripheral details.

Following this work, we intend to quantify coherence and diversity across the range of generated assets, and develop an automated game asset generation pipeline which may be approached from a quality-diversity angle [9].

This study is thus intended as a first step towards LLMs acting as a central coordination layer for automated game generation, streamlining the game development process while maintaining cohesive aesthetics.

References

[1] A. Liapis, G. N. Yannakakis, M. J. Nelson, M. Preuss, and R. Bidarra, "Orchestrating game generation," *IEEE Transactions on Games*, vol. 11, no. 1, pp. 48–68, 2019.

[2] M. Cook, S. Colton, and J. Gow, "The ANGELINA Videogame Design System—Part II," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 9, no. 3, 2017.

[3] R. Gallotta, G. Todd, M. Zammit, S. Earle, A. Liapis, J. Togelius, and G. N. Yannakakis, "Large Language Models and Games: A Survey and Roadmap," *arXiv preprint arXiv:2402.18659*, 2024.

[4] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[5] J. Dormans, "Adventures in level design: generating missions and spaces for action adventure games," in *Proceedings of the 2010 Workshop on Procedural Content Generation in Games*, 2010.

[6] A. Q. Jiang, A. Sablayrolles, A. Roux, A. Mensch, B. Savary, and C. Bamford et al., "Mixtral of experts," *arXiv preprint arXiv:2401.04088*, 2024.

[7] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, and J. Müller et al., "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis," *arXiv preprint arXiv:2307.01952*, 2023.

[8] L. Zhang, A. Rao, and M. Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models," *arXiv preprint arXiv:2302.05543*, 2023.

[9] M. Zammit, A. Liapis, and G. N. Yannakakis, "MAP-Elites with transverse assessment for multimodal problems in creative domains," in *Proceedings of the 6th International Conference on Computational Intelligence in Music, Sound, Art and Design (EvoMusArt)*, 2024.