One Spell Fits All: A Generative AI Game as a Tool for Research in AI Creativity and Sustainable Design

One Spell Fits All: A Generative AI Game as a Tool for Research in AI Creativity and Sustainable Design

Tom Tucek[1,*,†], Kseniia Harshina[1,*,†], Georgia Samaritaki[2,†] and Dipika Rajesh[3,†]

[1]University of Klagenfurt, Klagenfurt, 9020, Austria
[2]University of Amsterdam
[3]University of California, Santa Cruz

Abstract

This paper presents "One Spell Fits All", an AI-native game prototype where the player, playing as a witch, solves villagers' problems using magical conjurations. We show how, beyond being a standalone game, "One Spell Fits All" could serve as a research platform to explore several key areas in AI-driven and AI-native game design. These areas include AI creativity, user experience in predominantly AI-generated content, and the energy efficiency of locally running versus cloud-based AI models. By leveraging smaller, locally running generative AI models, including LLMs and diffusion models for image generation, the game dynamically generates and evaluates content without the need for external APIs or internet access, offering a sustainable and responsive gameplay experience. This paper explores the application of LLMs in narrative video games, outlines a game prototype's design and mechanics, and proposes future research opportunities that can be explored using the game as a platform.

Keywords

Generative AI, AI-driven Gameplay, Local AI Models, AI and Creativity, Player Experience, Procedural Content Generation

1. Introduction

In "One Spell Fits All" (OSFA), players take on the role of a witch, capable of conjuring an infinite variety of items using text-based input (see Figures 1 and 2). NPC-Villagers visit the witch with a wide range of problems, and the player is tasked with finding exactly one solution that solves all of their problems at once – using a single text input. The game uses locally running AI models to generate and evaluate game content, providing a sustainable alternative with lower latency compared to cloud-based AI services. Furthermore, we believe that this video game can serve as a research platform, as the game can provide an environment to investigate various aspects of generative AI, such as AI creativity, user experience for AI-generated content, and energy efficiency – all of which can be compared across different models or implementations. This paper discusses the game's integration of generative AI tools and outlines the potential research directions that can be pursued using this platform.

2. Related Work

Incorporating AI into interactive narratives has been a long-term objective in video game research [1]. This goal has seen significant advancements with the recent surge in generative AI tools, particularly large language models (LLMs). Furthermore, the use of LLMs in video games has been shown to increase the amount and the quality of possible player interactions [2, 3, 4, 5].

Significant advances for generative AI integrated into games have been shown with examples like "1001 Nights" [6, 7], a game which explores dynamic content generation by creating images of weapons based on the stories told by the player. The authors of the game also argue for the use of the term "AI-native" games, as opposed to AI-powered games, for video games that use generative models as an integral core feature. We adopt this term to describe OSFA.

The integration of AI in video games has been explored in literature, with foundational texts such as those by Yannakakis and Togelius [8] providing a comprehensive overview of how artificial intelligence can enhance game design through techniques like procedural content generation (PCG), player modeling, and dynamic difficulty adjustment. OSFA builds upon these concepts by applying local AI models to create a responsive and sustainable game environment. Elements of PCG [9] have formed the base of OSFA, to showcase how AI can autonomously create complex and engaging game content.

3. Game Design and Mechanics

The core gameplay loop of OSFA revolves around the player's text input. Taking on the role of a witch, players have to address the various needs of NPC-villagers by using magical conjurations. In each round, the system feeds a solution item from a list of predefined solutions to the LLM, which then reverse-engineers problems. These problems are then posed as requests from the villagers. Players have complete creative freedom regarding their input, which is funnelled into various models.

3.1. Gameplay Overview

In OSFA, an increasing number of villagers visit the player's hut, each with a specific problem that needs to be solved. The player must provide a solution through text input, to conjure an item that solves all of the present villagers' problems at once. The game progresses in turns, with each turn representing a new challenge through a new set of villagers. The player's success is measured by their ability to satisfy the villagers, which is reflected in the game's scoring system. Figures 1 and 2 showcase the in-game environment where the witch summons items based on player input (top-right) and the villagers voicing their problems.

EXAG '24: Experimental AI in Games Workshop at the AIIDE Conference, November 18, 2024, Lexington, USA
Tom.Tucek@aau.at (T. Tucek); Kseniia.Harshina@aau.at (K. Harshina); g.samaritaki@uva.nl (G. Samaritaki); dirajesh@ucsc.edu (D. Rajesh)
https://github.com/YenR/OneSpellFitsAll (T. Tucek)
0009-0001-1277-1473 (T. Tucek); 0009-0001-4491-4619 (K. Harshina); 0009-0006-8467-2374 (G. Samaritaki); 0009-0007-53 57-5560 (D. Rajesh)

Figure 1: Screenshot of OSFA, showing villagers and their problems. (The problems read: "Catching many small fish for dinner" and "Collecting fallen apples from [a] large tree".)

Figure 2: Screenshot of OSFA, showing the witch conjuring a chair and the villagers being satisfied with the solution. Note that this is not the solution to the problems in the previous screenshot.

3.2. Problem and Solution Generation

One of the innovative aspects of OSFA is the use of AI to dynamically generate problems and evaluate solutions at runtime. The game currently employs the *Mistral-7B-Instruct* model, via the *LLM for Unity* library [1] to generate unique problems equal to the number of present villagers, based on a provided keyword. This ensures that the challenges presented to the player are varied, contextually relevant, and have at least one correct solution for all problems. Once the player provides a solution, the game tests for validity and then uses the *all-MiniLM-L6-v2* [2] transformer model to evaluate the solution based on a similarity score to the original keyword. Future iterations of the game will try the use of LLMs for evaluation of the solution as well. The game then generates a visual representation of the solution, using a custom diffusion model [3] and *ComfyUI* [4], which is then presented to the player and the villagers.

3.3. Villager Interactions

The interactions between the player and the NPC-villagers are central to the game's narrative and mechanics. Villagers respond to the player's solutions based on the AI's evaluation, and their satisfaction is reflected in the game's scoring system. The scoring system, in turn, affects how many villagers come back to the player. These interactions are designed to create a dynamic and responsive gameplay loop, in which the player must think creatively and strategically to meet all of the villagers' needs.

[1]https://github.com/undreamai/LLMUnity
[2]https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2
[3]https://huggingface.co/megaaziib/aziibpixelmix
[4]https://github.com/comfyanonymous/ComfyUI

3.4. Artistic Elements

OSFA features a distinct pixel art style that complements its AI-driven gameplay. The game's art is a mix of free assets and AI-generated art, such as the sprite of the witch. The locally running model creates pixel art assets in real-time based on player input, which matches stylistically with the rest of the game. Furthermore, the background music and the theme song for the main menu have also been created using generative AI tools.

4. Technical Workflow

Figure 3 illustrates the workflow of OSFA, where various AI components generate and evaluate game content dynamically in the game's core loop.

The workflow is detailed as follows:

1. Villagers enter the witch's store, bringing their unique problems for the witch (the player) to solve.
2. The game uses an LLM to generate a problem for each villager. This AI model with custom instructions takes a sentence (in our case, a keyword) as input and produces problems, which are solvable by the keyword.
3. Each turn, the player must address all of the villagers' problems at once. The player inputs text – an item to be conjured by the witch to solve the problems.
4. The game then evaluates the player's solution using the *all-MiniLM-L6-v2* transformer model. This model checks the similarity between the player's input and the original keyword to determine if the problem has been correctly addressed. Note that this step could also be given to the LLM instead, which would then be able to evaluate the solution on a per-villager basis.
5. The game generates an image representing the solution. This image visually depicts the outcome of the player's input.
6. The game then calculates the player's score based on the success of their solution. The villagers give visual feedback on whether the solution was considered successful or not, leave the store, and a new loop begins.

5. AI Integration

OSFA uses multiple AI models running locally and at the same time to create a dynamic gameplay experience. The integration of these models is central to the game's ability to generate problems, evaluate solutions, and provide visual feedback. This section gives more details on how various AI components are implemented and interact within the game. Notably, the game's architecture was designed with modularity in mind – which means that each model can be swapped out separately.

The *Mistral-7B-Instruct* model, used as the LLM within the game, is responsible for generating the unique problems that each villager presents to the player. The model operates locally on the player's computer, limiting energy consumption and eliminating the need for external API calls. The use

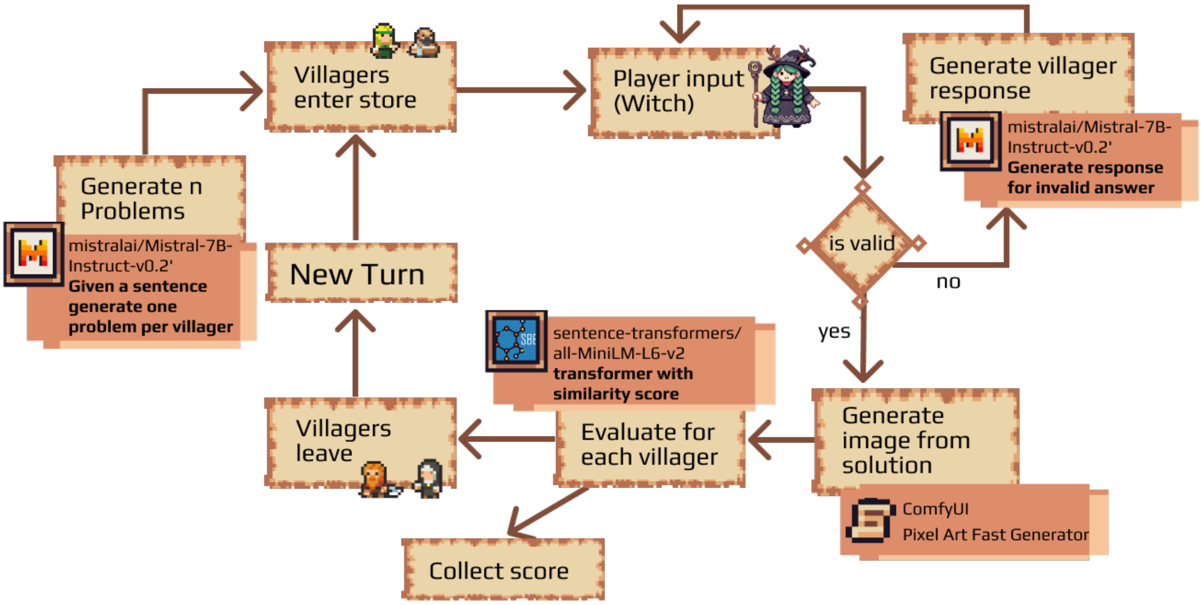

Figure 3: Workflow diagram showing game loop for OSFA and the interaction between AI components and user input in the game.

of such a local implementation can also reduce the latency between player input and feedback, as well as improve data privacy concerns, as no data is sent to third parties.

To evaluate the player's solutions, OSFA integrates the *all-MiniLM-L6-v2* transformer model. After the player provides a solution to a villager's problem, the model calculates a similarity score compared to a 'perfect' solution, thus determining whether the player's input sufficiently addresses the problem. Currently, all villagers' problems are evaluated at once, but future implementations will aim to evaluate them on a case-by-case basis (e.g., by using an LLM instead of a transformer), thus allowing for partially correct answers as well.

Based on the player's input, the conjured items are generated using a pixel-art diffusion model and a custom *ComfyUI workflow*. This tool is integrated into the game's AI pipeline, allowing it to dynamically create pixel art assets based on the input in real time.

The AI models in OSFA are implemented within a framework that supports real-time interaction and processing. The models run locally on the player's machine and the game was made using the Unity Game Engine, using scripts, plugins, and libraries to control the AI models. By running the AI components locally, the game minimizes latency and increases security through the lack of network communication, although the player's hardware specification can still cause delays.

6. User Experience and Feedback

Initial testing indicates that players appreciate the dynamic problem-solving elements, though challenges such as model latency were encountered and mitigated through optimization techniques. The game was developed and showcased during a game jam event, where it won second place. More than ten people at a time engaged with the game, often collaborating to guess and influence the player's decisions. This collective engagement fostered curiosity and a fun, cooperative atmosphere. Players found it challenging to discover the correct solutions but remained persistent, continually attempting new ideas. When the villagers left satisfied with a solution, it elicited cheers and a sense of accomplishment among the group. Participants particularly enjoyed the generated images, highlighting them as a standout feature. However, some areas of improvement were noted. There was ambiguity when players felt their solutions were appropriate but unacknowledged by the system. Enhancing the feedback mechanism to provide clearer guidance when solutions are rejected could significantly improve user experience.

7. Future Work

Future work will focus on evaluating and enhancing the AI-driven creativity within the game, assessing user experience and satisfaction through comprehensive playtesting, and analyzing the cost-benefit ratio of using local AI models compared to cloud-based solutions. We aim to ensure an engaging and satisfying player experience while at the same time promoting sustainable AI practices in game development.

7.1. AI Creativity

Creativity, especially in the context of generative AI, is a rich and interdisciplinary research field [10, 11, 12]. As OSFA uses an LLM to generate villagers' wishes and requests based on specific keywords, we believe that it opens up new possibilities for exploring AI creativity. The ability of AI to generate novel and varied problems from similar inputs is understood to be a form of creativity, presenting several interesting research questions. The first step could be defining and quantifying AI creativity in the context of the game prototype. One approach could be to develop a *creativity score* to evaluate the LLM's performance in generating unique and contextually appropriate problems. This score would be based on the similarity of problems generated by the AI across various playthroughs and different versions of the game.

To explore this, data on AI-generated problems needs to be collected across multiple sessions and then compared against several types of solutions: those manually crafted by human designers, those generated solely by the LLM, and a mix of human and AI-generated content. This comparison could reveal how closely AI outputs align with expected creative standards, as well as how AI can complement human creativity.

In addition to a creativity score, other metrics can help to better understand and measure AI creativity. *Novelty* can assess how often the AI produces entirely new problems not seen in previous playthroughs. *Diversity* measures the range of different problem types generated in response to similar keywords within a single playthrough. *Consistency* can evaluate the AI's ability to maintain high-quality, contextually relevant problems across different scenarios. *Adaptability* assesses how well the AI adjusts its problem generation to varying game contexts or player actions, ensuring that the problems remain appropriate and challenging.

This leads to another question – how do different LLM models compare in terms of their creative outputs? By systematically comparing these metrics across various LLMs, we could gain insights into which models are better suited for generating creative content in games. This could also inform the future development of AI systems designed specifically for creative tasks. Developing these metrics and comparing different models should deepen our understanding of how AI can be used to generate novel and engaging content. This research could have broader implications for AI applications in creative industries, such as game design, storytelling, and digital art.

7.2. User Experience and Satisfaction

Understanding user experience (UX) and satisfaction can help in evaluating the success of OSFA, especially given its integration of AI-generated content. This section explores methods to assess how players interact with and perceive the game, particularly in the context of its AI elements. One approach to evaluating UX in OSFA is heuristic evaluation. By having experts review the game based on established usability principles, it's possible to identify both strengths and potential issues within the game. Key considerations include ensuring that AI-generated content maintains consistency and adheres to the game's internal rules.

Beyond heuristic evaluation, another important aspect is understanding how players' awareness of AI-generated content influences their enjoyment. It is worth investigating whether players perceive the AI elements (referring to both the pre-generated content such as music and visuals, as well as dynamically generated content) as positive additions to the game or if knowing that most of the content is AI-generated affects their perception of the game's creativity and quality. For instance, surveys or interviews could be conducted to explore whether this awareness impacts players' overall satisfaction and engagement and how AI-driven design challenges their expectations. In summary, understanding user experience and satisfaction in a predominantly AI-generated game like OSFA can help answer other important research questions regarding players' perceptions of AI content.

7.3. Energy Efficiency of Local AI Models vs. Cloud-Based Solutions

As AI-native games like OSFA continue to develop, an important question arises: How does the utilization of local AI models compare to cloud-based solutions in terms of energy efficiency? This question is particularly relevant for our game, which operates AI models locally. While it is generally accepted that running smaller LLMs locally would be faster and more power efficient than remote calls to larger cloud-based models (such as GPT-5), this claim still warrants further investigation, especially in terms of broader scalability and environmental impact. A more detailed comparative analysis could explore the trade-offs between local and cloud-based solutions, considering the quality of the generated content, latency, and also the total energy consumption involved. This includes power requirements for both local machines and cloud-based servers, as well as the energy costs of data transmission between clients and remote servers. The goal would be to determine which approach offers greater energy efficiency, considering factors such as scalability, environmental impact, and the trade-offs between performance, quality, latency, and energy use.

Understanding these differences can have significant implications for game development. If local AI models prove to be more energy-efficient while guaranteeing sufficient quality of experience, they could become the preferred choice for sustainable game design. However, if their quality remains lacking, and cloud-based solutions offer better efficiency, this will influence how AI resources are deployed in future projects as well. Exploring the energy efficiency of AI models is important for many reasons – not only for improving game performance but also for advancing sustainability in the AI and video game industry. By comparing the energy use of local AI models with server-based and cloud-based solutions, this research could guide future developments in AI-driven game design, promoting more environmentally conscious practices.

8. Conclusion

OSFA demonstrates the potential of locally-run AI models in game development, offering both an engaging gameplay experience and a versatile platform for research. By reducing reliance on cloud-based services, the game opens up new avenues for studying AI creativity, UX, and energy efficiency. This paper outlines the game's design and mechanics, while also proposing several research directions that can be explored using the game as a platform, thus contributing to the broader discourse on AI in games and digital creativity.

References

[1] M. O. Riedl, V. Bulitko, Interactive narrative: An intelligent systems approach, Ai Magazine 34 (2013) 67–67.

[2] T. Ashby, B. K. Webb, G. Knapp, J. Searle, N. Fulda, Personalized quest and dialogue generation in role-playing games: A knowledge graph-and language model-based approach, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–20.

[3] S. Värtinen, P. Hämäläinen, C. Guckelsberger, Generating role-playing game quests with GPT language models, IEEE transactions on games (2022).
[4] Y. Wang, Q. Zhou, D. Ledo, Storyverse: Towards co-authoring dynamic plot with LLM-based character simulation via narrative planning, arXiv preprint arXiv:2405.13042 (2024).
[5] Latitude, AI Dungeon, https://play.aidungeon.io/, 2019. Accessed: 2024-09-27.
[6] Y. Sun, Z. Li, K. Fang, C. H. Lee, A. Asadipour, Language as reality: A co-creative storytelling game experience in 1001 nights using generative AI, Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment 19 (2023) 425–434. URL: https://ojs.aaai.org/index.php/AIIDE/article/view/27539. doi:10.1609/aiide.v19i1.27539.
[7] Ada Eden, 1001 Nights AI: Discover generative stories, 2024. URL: https://www.1001nights.ai/, accessed: 2024-08-30.
[8] G. N. Yannakakis, J. Togelius, Artificial intelligence and games, Springer, 2018.
[9] D. Gravina, A. Khalifa, A. Liapis, J. Togelius, G. N. Yannakakis, Procedural content generation through quality diversity, in: 2019 IEEE Conference on Games (CoG), IEEE, 2019, pp. 1–8.
[10] J. Rowe, D. Partridge, Creativity: A survey of AI approaches, Artificial Intelligence Review 7 (1993) 43–70.
[11] J. H. Roosa Wingström, R. Lundman, Redefining creativity in the era of AI? perspectives of computer scientists and new media artists, Creativity Research Journal 36 (2024) 177–193. doi:10.1080/10400419.2022.2107850.
[12] L.-C. Lu, S.-J. Chen, T.-M. Pai, C.-H. Yu, H.-y. Lee, S.-H. Sun, LLM discussion: Enhancing the creativity of large language models via discussion framework and role-play, arXiv preprint arXiv:2405.06373 (2024).