

Explorando a Aplicação de Modelos de Linguagem de Grande Escala (LLMs) para o Design de Game Stories

Exploring the Application of Large Language Models (LLMs) to Design Game Stories

Victor Travassos Sarinho¹

¹ Laboratório de Entretenimento Digital Aplicado (LEnDA)
Universidade Estadual de Feira de Santana (UEFS)
Feira de Santana, BA – Brasil

vsarinho@uefs.br

Abstract. Introduction: *In the context of digital games, Large Language Models (LLMs) have shown the potential to significantly transform content generation, as well as influence tools, development processes, and research approaches in the field of game design. Objective:* This study aims to investigate the potential of using structured prompts applied to LLMs in the digital game design process. **Methodology or Steps:** *Initially, structured game stories will be generated based on a common theme using four different LLMs that are freely available. Next, the generated game design elements will be evaluated according to criterias such as creativity, playability, correctness, and usefulness. Results:* The results indicate that LLMs are capable of generating structured game design elements with satisfactory quality, according to the defined evaluation criteria, thus highlighting the potential of these narrative structures to support automated processes of conception and prototyping of desired digital games.

Keywords Game Design, Large Language Models (LLMs), Game Stories.

Resumo. Introdução: *No contexto de jogos digitais, os Modelos de Linguagem de Grande Escala (LLMs) apresentam a capacidade de transformar significativamente a geração de conteúdo, bem como impactar ferramentas, processos de desenvolvimento e abordagens de pesquisa na área de design de jogos. Objetivo:* Este trabalho tem como objetivo investigar o potencial do uso de prompts estruturados aplicados a LLMs no processo de design de jogos digitais. **Metodologia ou Etapas:** *Inicialmente, serão geradas game stories estruturadas a partir de uma temática comum, utilizando quatro LLMs distintas disponíveis gratuitamente. Em seguida, os elementos de design produzidos serão avaliados com base em critérios como criatividade, jogabilidade, corretude e utilidade. Resultados:* Os resultados indicam que os LLMs são capazes de gerar elementos estruturados de design de jogos com qualidade satisfatória, de acordo com os critérios de avaliação definidos, evidenciando assim o potencial dessas estruturas narrativas para apoiar processos automatizados de concepção e prototipagem de jogos digitais desejados.

Palavras-Chave Design de Jogos, Modelos de Linguagem de Grande Escala (LLMs), Game Stories.

1. Introdução

Os Modelos de Linguagem de Grande Escala (LLMs) revolucionaram o cenário científico, transformando profundamente o campo do processamento de linguagem natural e a interação entre humanos e computadores como um todo [Lanzi e Loiacono 2023]. Como modelos computacionais capazes de compreender e gerar linguagem humana em larga escala, os LLMs demonstram um notável potencial para responder a perguntas complexas e realizar tarefas criativas desafiadoras. Entre essas tarefas destacam-se a geração de código, o desenvolvimento de aplicações voltadas à resolução de problemas, a escrita de narrativas e obras artísticas, entre outros tipos de produção textual e criativa [Lanzi e Loiacono 2023].

Segundo [Radford et al. 2019], LLMs podem atuar como ferramentas de apoio no aprimoramento e na ampliação do processo manual de design de jogos. Esses modelos têm o potencial de transformar profundamente tanto a geração quanto a cocriação de conteúdo em videogames, além de impactar significativamente ferramentas, processos de desenvolvimento e abordagens de pesquisa na área de jogos [Sweetser 2024]. Estudos anteriores já exploraram diversas aplicações dos LLMs nesse contexto, como a geração de diálogos entre jogadores e personagens não jogáveis [Volum et al. 2022], a facilitação da criação de novos elementos no mundo do jogo impulsionada pelas ações dos jogadores [Huang e Sun 2023], e o auxílio na descoberta de novos caminhos narrativos em jogos baseados em texto [Alavi et al. 2024], entre outras possibilidades [Sweetser 2024].

No entanto, para que os LLMs sejam plenamente utilizados em seu potencial, é fundamental que as instruções sejam cuidadosamente formuladas, uma vez que a eficácia desses modelos está diretamente relacionada à qualidade dos *prompts* empregados [Marvin et al. 2023]. Assim, a elaboração de prompts bem estruturados é crucial para a obtenção de respostas adequadas e relevantes, ao passo que instruções mal formuladas tendem a resultar em saídas insatisfatórias para os seus respectivos usuários [Heston e Khun 2023].

Diferentes iniciativas têm sido desenvolvidas com o objetivo de padronizar a produção de requisitos no design de jogos, a exemplo do uso de *User Stories* para representar elementos e interações no contexto do design de jogos [Sarinho 2020]. Tal abordagem, ao empregar uma linguagem natural estruturada, possibilita uma orientação inicial simplificada do trabalho de design na representação de características de jogos desejados, bem como a produção de resultados estratificados capazes de serem usados em abordagens automatizadas de manipulação de requisitos de jogos em si, contribuindo assim para a melhoria da comunicação e da produtividade de designers, desenvolvedores e demais partes interessadas nos jogos a serem produzidos.

Nesse contexto, este trabalho busca investigar o potencial de utilização de prompts estruturados aplicados a LLMs no processo de design de jogos digitais. Para isso, são apresentados os resultados da geração de *game stories* estruturadas a partir de quatro LLMs distintas, todas de acesso gratuito. Como contribuição deste trabalho, espera-se validar a viabilidade da produção automática de elementos estruturados de design de jogos com base em um modelo de prompt previamente definido, a partir da análise de critérios como criatividade, jogabilidade, corretude e utilidade, aplicados na avaliação dos elementos de design automaticamente gerados, oferecendo assim insights práticos para o uso de LLMs em processos automatizados de prototipagem e concepção de jogos.

2. Trabalhos Relacionados

Conforme pesquisa realizada no *Google Scholar*¹, alguns estudos recentes publicados nos últimos 5 anos têm explorado a aplicação prática de LLMs no uso de prompts estruturados para a modelagem de requisitos de jogos digitais. Como exemplo, [Lanzi e Loiacono 2023] propõem um framework em que os usuários representam ideias de design por meio de texto em linguagem natural livre (*free-form*), sendo todas as operações de design sendo realizadas através de consultas a um LLM. Nesse contexto, os prompts utilizados para gerar ideias iniciais de jogos consistem em: (i) um briefing de design, estruturado com objetivos e restrições, e (ii) uma instrução para que o LLM atue como um designer de jogos, criando um projeto original com base nas informações fornecidas. Essa abordagem permite o refinamento progressivo das respostas, direcionando o LLM a atender e gerar requisitos específicos do jogo modelado.

[Saito et al. 2023] relatam um estudo de caso em uma atividade do Academy Camp, um acampamento científico sem fins lucrativos, na qual crianças utilizaram um LLM, neste caso o ChatGPT², para criar seus próprios jogos de RPG de mesa (*Tabletop RPGs*) como jogos sérios sobre temas sociais de seu interesse. O modelo de prompt aplicado nesse estudo incluía restrições genéricas sobre o jogo, tais como: o ChatGPT é o mestre do jogo, NPCs possuem seus próprios interesses e objetivos, dentre outros, além de um espaço específico para que os usuários descrevessem as características desejadas, tais como: o ponto de partida da narrativa, a jornada do jogador e o objetivo a ser alcançado.

Por sua vez, [Junior et al. 2023] investiga o uso do ChatGPT na criação de jogos de tabuleiro educacionais, concebidos como objetos de pensamento (*objects-to-think-with*). O estudo propõe uma orientação para educadores em três fases do design: ideação (com sugestões de temas e mecânicas alinhadas aos objetivos de aprendizagem), customização (com modelos para jogos personalizados) e avaliação (com sugestões para aprimoramento de componentes e mecânicas). Para cada fase do design, são utilizados prompts estruturados para a realização de perguntas sobre o jogo em questão, tais como: “*Proponha conceitos para o desenvolvimento de um jogo de tabuleiro educacional para [definir o público-alvo: adultos, jovens, etc.], com foco em promover e reforçar habilidades essenciais de [tema ou objetivo curricular, como alfabetização financeira]*”, os quais permitem refinar iterativamente as respostas obtidas.

A partir da análise desses trabalhos, onde apenas respostas geradas pelo ChatGPT foram consideradas, observa-se que [Lanzi e Loiacono 2023] empregam uma abordagem de refinamento de respostas a partir de *prompts genéricos* complementados por características iniciais do jogo, os quais podem gerar resultados imprecisos e ambíguos. Já [Saito et al. 2023] optam por restringir o escopo das respostas do LLM, obtendo *descrições genéricas e superficiais*, porém centradas na ideia geral dos jogos desejados. Por fim, [Junior et al. 2023] realizam *múltiplos refinamentos* ao longo do processo de interação com o LLM, onde foi possível identificar algumas limitações, tais como *geração de detalhes imprecisos*, propostas de *regras contraintuitivas* e *interpretações equivocadas* de feedbacks, o que resultou em dinâmicas indesejadas nos jogos concebidos.

¹<https://scholar.google.com/>

²<https://openai.com/index/chatgpt>

3. Metodologia

Para a realização do trabalho proposto, definiu-se um modelo de prompt voltado para a geração de *game stories* estruturadas. Trata-se de um modelo fundamentado em um estudo que propõe estruturas narrativas e lógicas baseadas na representação de *User Stories* voltadas para o design de jogos [Sarinho 2020]. Nesse contexto, o prompt foi elaborado com base em restrições iniciais que englobam aspectos relacionados a espaços, atores, itens, interações, regras e desafios do jogo proposto, conforme ilustrado na Figura 1.

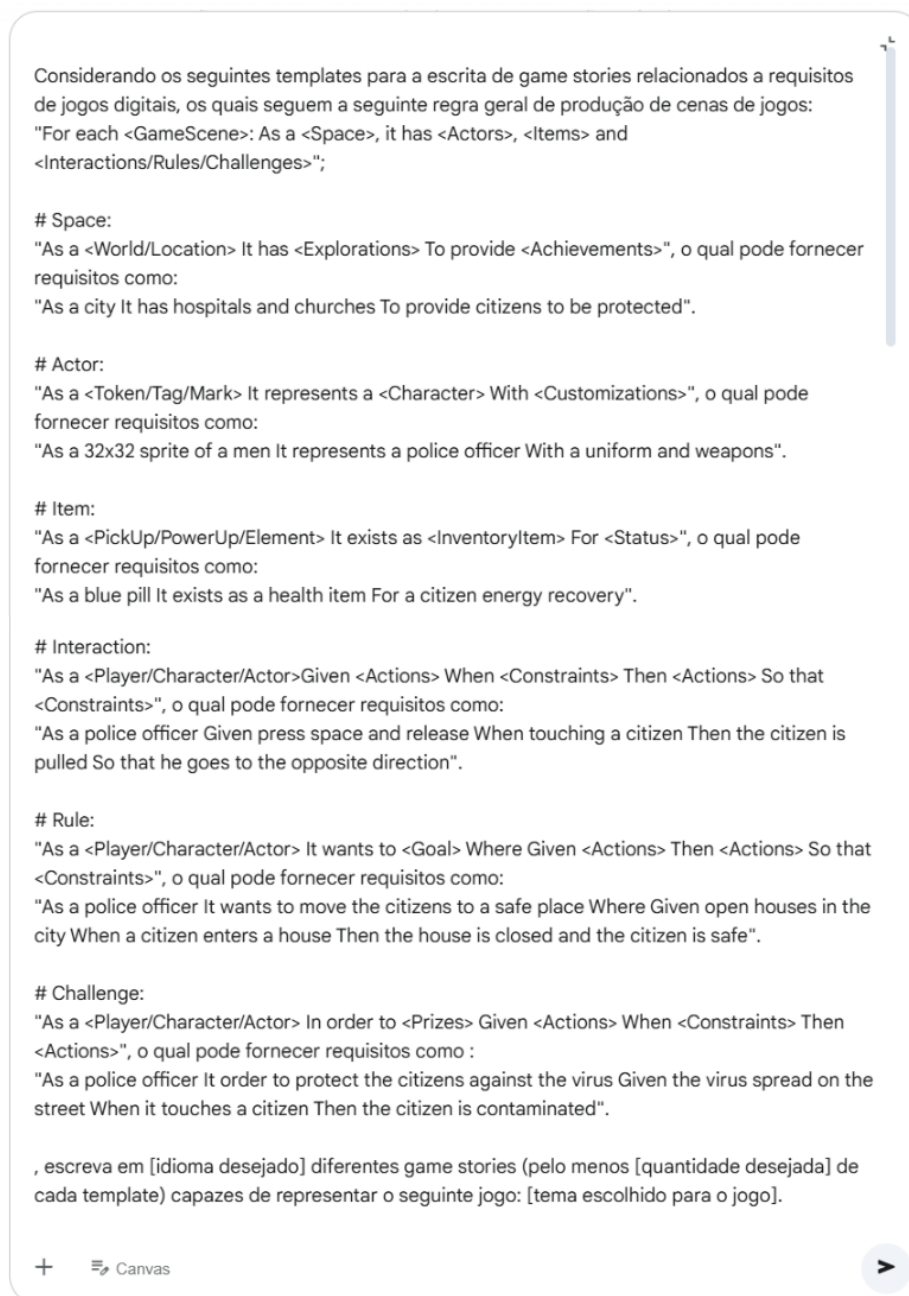


Figura 1. Modelo de prompt para a geração de game stories estruturadas.

Com o modelo de prompt definido, selecionou-se LLMs de acesso gratuito para a

realização das respectivas consultas. Assim, foram utilizados os seguintes LLMs devido à sua disponibilidade pública e capacidade de geração textual coerente para o design de jogos: ChatGPT (versão GPT-4-turbo), Gemini³ (versão 2.0 Flash), DeepSeek⁴ (versão V3) e Llama⁵ (versão 4). Cada prompt aplicado nos LLM escolhidos seguiram uma temática comum — neste caso, uma aventura textual infantil inspirada no universo do jogo Pac-Man —, sendo submetido uma única vez a cada LLM, de modo a trabalhar com resultados originais, espontâneos e não enviesados por múltiplas iterações ou ajustes sucessivos.

Para verificar e validar as game stories produzidas, foram analisados aspectos relacionados à criatividade, jogabilidade, corretude e utilidade dos elementos gerados. Para isso, elaborou-se um conjunto de perguntas avaliativas, conforme demonstradas a seguir, que foram analisadas com base nas respostas geradas pelas LLMs utilizadas:

- Os elementos de espaço, atores e itens obtidos apresentam originalidade e consistência temática?
- As interações de jogo descritas são claras, plausíveis e jogáveis?
- A narrativa com base nas interações proposta apresenta originalidade e consistência com os elementos do jogo?
- Há coerência lógica entre os elementos/eventos narrativos e os objetivos do jogador?
- As metas e desafios propostos são possíveis de serem atingidos e conquistados pelo jogador?
- Os jogos produzidos conseguem ter um início, meio e fim bem definidos?
- Os elementos apresentados oferecem suporte ao desenvolvimento ou prototipagem de um jogo digital funcional?

Cada pergunta foi respondida utilizando uma escala de concordância de 1 (discordo fortemente) a 7 (concordo fortemente), conforme o modelo de questionário USE [Lund 2001]. Além disso, foi realizada uma análise qualitativa dos pontos fortes e fracos de cada jogo gerado, permitindo uma avaliação mais abrangente dos resultados obtidos por cada LLM.

Por fim, com base nos dados quantitativos e qualitativos coletados, foi conduzida uma análise crítica das potencialidades e limitações de cada LLM na geração de game stories estruturadas. Essa avaliação considerou aspectos como padrões narrativos, consistência temática e eventuais incoerências lógicas presentes nos elementos estruturados produzidos, visando identificar contribuições e desafios para o uso de LLMs no apoio ao design automatizado de jogos digitais.

4. Resultados Obtidos

Para a aplicação do prompt estruturado nas LLMs selecionadas, estabeleceu-se que cada modelo deveria gerar, em português, game stories voltadas para uma aventura textual infantil inspirada no jogo Pac-Man. A exigência era que fossem apresentadas pelo menos 5 game stories para cada um dos seguintes aspectos: espaços, atores, itens, interações,

³<https://gemini.google.com>

⁴<https://www.deepseek.com>

⁵<https://www.llama.com/>

regras e desafios, totalizando no mínimo 30 game stories por resposta. A definição dessa quantidade mínima ocorreu porque respostas com menos de 5 stories por categoria apresentaram pouca diversidade de conteúdo, enquanto que respostas com mais de 5 stories resultaram, em alguns modelos, na repetição de temas e conteúdos narrativos.

Assim, considerando as respostas do ChatGPT, do DeepSeek e do Llama, estes apresentaram 5 game stories para cada um dos tipos relacionados, totalizando exatamente 30 game stories para cada resposta. Já o Gemini apresentou uma resposta com um total de 48 game stories, organizadas em 5 *game scenes*, descrevendo apenas uma game story de espaço para cada game scene apresentada. A Tabela 1 ilustra algumas game stories produzidas por cada LLM conforme os tipos de game stories indicados.

Tabela 1. Exemplos de game stories produzidas pelos LLM avaliados.

	Llama	ChatGPT	DeepSeek	Gemini
Espaço	Como um labirinto Ele tem caminhos e esquinas Para fornecer um desafio para Pac-Man	Como um castelo de luzes coloridas, ele possui salões brilhantes e passagens ocultas para proporcionar desafios e encontros inesperados	Como um labirinto colorido, ele tem túneis e corredores Para fornecer caminhos cheios de desafios e doces para coletar	Como um labirinto mágico, ele tem caminhos coloridos e surpresas escondidas para proporcionar uma aventura divertida
Ator	Como um fantasma vermelho Ele representa Blinky Com olhos brancos e uma expressão zangada	Como um ícone de moedinha dourada, ele representa a Moeduca, personagem que guia os jogadores pelos caminhos secretos	Como um power-up azul, ele representa a pilula energética Com um brilho pulsante para indicar seu efeito	Como 'Fantasma-Risadinha', ele representa um personagem brincalhão que adora brincar de esconde-esconde
Item	Como uma pastilha Ele existe como um ponto Para aumentar a pontuação de Pac-Man	Como uma pilula dourada, ela existe como item de energia para restaurar a coragem de Pac-Guto	Como um escudo prateado, ele existe como um item de defesa Para proteger o Pac-Man por alguns segundos	Como 'Portal Arco-Íris', ele existe como um item de teletransporte para levar o Pac-Amigo para diferentes partes do labirinto
Interação	Como Blinky Dado perseguir Pac-Man Quando Pac-Man está perto Então Blinky se aproxima Tal que pode capturar Pac-Man	Como o Pac-Guto, dado pressionar a seta para cima, quando em frente a uma parede secreta, então atravessa a parede, para que encontre um item escondido	Como uma cereja, Dado aparecer em um canto do mapa Quando o Pac-Man a coleta Então ela some e dá 100 pontos Para recompensar o jogador	Como Pac-Amigo, dado comer um Pirulito Colorido, quando a energia doce está baixa, então a energia doce aumenta, para que Pac-Amigo possa continuar explorando
Regra	Como Pac-Man Ele quer coletar todos os pontos Onde Dado se mover pelo labirinto Quando não há obstáculos Então Pac-Man coleta pontos Tal que pode concluir o nível	Como Moranguinha, ela quer ajudar Pac-Guto a escapar, onde dado encontrar um Fantasmico, então usa o feitiço de confete, para que o inimigo fuja	Como o jogo, ele quer oferecer desafios Onde Dado o Pac-Man coletar todas as frutas Então um novo labirinto é gerado Para continuar a aventura	Como Pac-Amigo, ele quer coletar todas as Estrelinhas Brilhantes, onde dado mover pelo labirinto, então as estrelinhas são coletadas, para que o labirinto fique cheio de alegria
Desafio	Como um fantasma Para tocar Pac-Man Dado se mover pelo labirinto Quando Pac-Man não está invencível Então o fantasma toca Pac-Man Tal que Pac-Man perde uma vida	Como Estelina, para proteger os amigos, dado ativar a estrela mágica, quando três Fantasmicos estiverem próximos, então paralisa todos por 5 segundos	Como o sistema de recompensas, a fim de motivar o jogador Dado pegar uma fruta rara Quando ela aparece aleatoriamente Então ele ganha uma vida extra	Como Pac-Amigo, para pegar todas as estrelas cadentes, dado voar pelo céu, quando Fantasma-Nuvem tenta brincar de esconde-esconde, então Pac-Amigo deve usar seu Telescópio Mágico para encontrar as estrelas escondidas

Com relação ao conteúdo das respostas produzidas nas game stories, segue abaixo uma análise qualitativa dos resultados obtidos:

- **Espaços:** O Llama adotou uma abordagem clássica e funcional, descrevendo cenários como labirintos e cidades, mas sem muita originalidade. O ChatGPT destacou-se pela criatividade, criando locais como ilhas flutuantes e bosques encantados. O DeepSeek equilibrou cenários tradicionais com propostas lúdicas como castelos de doces e florestas mágicas. Já o Gemini trouxe ambientes bem personalizados e temáticos, como céus estrelados e rios de bolhas, integrando melhor narrativa e ambientação.

- **Atores:** O Llama manteve personagens tradicionais e descrições diretas, a exemplo da game story: *“Como um fantasma vermelho Ele representa Blinky Com olhos brancos e uma expressão zangada”*. O ChatGPT criou personagens novos e carismáticos, como Pac-Guto e Moeduca, atribuindo funções narrativas, conforma a seguinte game story de exemplo: *“Como um ícone de moedinha dourada, ele representa a Moeduca, personagem que guia os jogadores pelos caminhos secretos”*. O DeepSeek combinou personagens clássicos com pequenas inovações, a exemplo da game story: *“Como um NPC simpático, ele representa um velho sábio Com um chapéu e um mapa para ajudar o jogador”*. Já o Gemini adaptou Pac-Man e fantasmas para diferentes cenários fictícios, a exemplo do Pac-Amigo, um construtor na Montanha dos Brinquedos.
- **Itens:** O Llama descreveu itens tradicionais com funções básicas, como “pastilha para pontuação”. O ChatGPT inventou objetos narrativos como “pílula dourada” e “relógio encantado”. O DeepSeek trouxe itens clássicos com leves variações funcionais a exemplo da “cereja vermelha” e da “pílula azul reluzente”. Já o Gemini criou itens integrados aos cenários, como “Pirulito Colorido” para aumentar energia doce e “Bolha Gigante” para flutuar.
- **Interações:** O Llama seguiu o padrão clássico, seguindo uma lógica simples de condição → ação → resultado, a exemplo da game story: *“Como Pac-Man Dado pressionar as teclas de direção Quando não há obstáculos Então Pac-Man se move Tal que ele pode coletar pontos”*. O ChatGPT inovou com interações narrativas, a exemplo da game story: *“Como a Moranguinha, dado pressionar espaço, quando próxima a um Fantasmico, então usa um feitiço de confete, para que o Fantasmico fique tonto”*. O DeepSeek trouxe variações funcionais, como túneis e power-ups, a exemplo da game story: *“Como um túnel secreto, Dado o Pac-Man entrar por um lado Quando ele chega ao final Então ele sai do outro lado Para criar um atalho”*. Por fim, o Gemini personalizou interações conforme o cenário, a exemplo da game story: *“Como Pac-Amigo, dado comer um Pirulito Colorido, quando a energia doce está baixa, então a energia doce aumenta, para que Pac-Amigo possa continuar explorando”*.
- **Regras:** O Llama manteve as regras originais, a exemplo da game story: *“Como Pac-Man Ele quer coletar todos os pontos Onde Dado se mover pelo labirinto Quando não há obstáculos Então Pac-Man coleta pontos Tal que pode concluir o nível”*. O ChatGPT personalizou regras, a exemplo da game story: *“Como Estelina, ela quer manter a luz acesa no labirinto, onde dado um Fantasmico apagar a luz, então usa a estrela mágica, para que a luz retorne”*. O DeepSeek mesclou regras clássicas com variações, incluindo geração de labirintos e novos power-ups, a exemplo da game story: *“Como o power-up, ele quer ajudar o jogador Onde Dado o Pac-Man pegá-lo Quando os fantasmas estão próximos Então eles ficam vulneráveis Para serem comidos”*. Por fim, o Gemini criou regras específicas por cenário, a exemplo da game story: *“Como Pac-Amigo, ele quer coletar todas as Estrelinhas Brilhantes, onde dado mover pelo labirinto, então as estrelinhas são coletadas, para que o labirinto fique cheio de alegria”*.
- **Desafios:** O Llama repetiu não só padrão clássico como praticamente aplicou os mesmos textos das games stories de regras, a exemplo da game story: *“Como Pac-Man Para coletar todos os pontos Dado se mover pelo labirinto Quando não há obstáculos Então Pac-Man coleta pontos Tal que pode concluir o nível”*.

O ChatGPT diversificou, adicionando desafios de tempo, proteção e missões bônus, a exemplo da game story: *“Como jogador, para acessar a fase bônus, dado coletar todas as frutas vermelhas, quando sem perder nenhuma vida, então destrava a fase secreta”*. O DeepSeek variou cenários e contextos, incluindo túneis, cercos e recompensas aleatórias, a exemplo da game story: *“Como os fantasmas, a fim de dificultar o jogo Dado perseguir o Pac-Man Quando ele está sem power-ups Então eles devem cercá-lo”*. Já o Gemini criou desafios narrativos integrados, a exemplo da game story: *“Como Pac-Amigo, para pegar todas as estrelas cadentes, dado voar pelo céu, quando Fantasma-Nuvem tenta brincar de esconde-esconde, então Pac-Amigo deve usar seu Telescópio Mágico para encontrar as estrelas escondidas”*.

Com relação as perguntas avaliativas elaboradas para verificar e validar aspectos relacionados à criatividade, jogabilidade, corretude e utilidade das respostas geradas pelos LLMs, tem-se na Tabela 2 as pontuações em uma escala de 1 a 7 para as mesmas, juntamente com as respectivas justificativas para cada nota aplicada.

Tabela 2. Resultado da avaliação dos aspectos relacionados à criatividade, jogabilidade, corretude e utilidade das respostas geradas pelos LLMs.

Critério	Llama	ChatGPT	DeepSeek	Gemini
1- Originalidade e consistência temática em Espaço, Atores e Itens <i>Justificativa:</i> Llama com respostas básicas e tradicionais ao clássico Pac-Man. DeepSeek também mantém o clássico porém com pequenas melhorias. ChatGPT e Gemini expandem criativamente com temas e itens bem integrados (Gemini se destaca pelo universo narrativo lúdico).	3	6	5	7
2- Clareza, plausibilidade e jogabilidade nas Interações <i>Justificativa:</i> Llama cumpre o básico, mas sem variedade. DeepSeek com respostas funcionais e coerentes. ChatGPT e Gemini criam interações bem definidas, jogáveis e interessantes.	4	7	6	7
3- Originalidade e consistência narrativa baseada nas Interações <i>Justificativa:</i> Llama com respostas previsíveis. DeepSeek melhora um pouco em relação ao Llama, mas ChatGPT e Gemini criam narrativas compatíveis e criativas, com ações integradas ao contexto.	3	7	5	6
4- Coerência lógica entre elementos/eventos narrativos e os objetivos do <i>Justificativa:</i> Todos mantêm coerência básica. ChatGPT e Gemini enriquecem com objetivos paralelos e conexões claras entre ações e regras do jogo.	5	7	6	6
5- Viabilidade de metas e desafios serem atingidos/conquistados pelo jogador <i>Justificativa:</i> Todos propõem desafios possíveis. Llama é básico, DeepSeek sólido, ChatGPT e Gemini oferecem objetivos acessíveis com boas variações e possibilidades de vitória.	6	7	7	7
6- Definição de início, meio e fim nos jogos propostos <i>Justificativa:</i> Llama não deixa clara a estrutura narrativa completa. DeepSeek sugere progressão do jogo nas respostas. ChatGPT e Gemini estruturam aventuras com começo, desenvolvimento e desfecho associados.	4	7	6	7
7- Suporte ao desenvolvimento/prototipagem de um jogo digital funcional <i>Justificativa:</i> Llama atende a demanda para protótipos que seguem o clássico Pac-Man. DeepSeek também. ChatGPT e Gemini apresentam material rico e aplicável em prototipagem de jogos funcionais, especialmente narrativos e educativos.	5	7	6	7

Por fim, com relação aos pontos fortes e pontos fracos identificados em cada LLM avaliado, a Tabela 3 apresenta um resumo consolidado dessas características, com base nas respostas obtidas para os diferentes tipos de game stories. De modo geral, a análise revelou que o ChatGPT e o Gemini se destacaram pela criatividade narrativa, adaptação ao tom infantil e pela capacidade de contextualizar elementos do jogo de forma lúdica

e acessível. Por outro lado, Llama e DeepSeek apresentaram maior rigidez estrutural e aderência à lógica clássica de Pac-Man, demonstrando menor flexibilidade criativa e inventividade na adaptação de enredos. O ChatGPT mostrou-se eficiente em equilibrar storytelling e estrutura formal de game stories, ainda que, em algumas ocasiões, tenha priorizado a narrativa em detrimento de aspectos de mecânica ou regras. Já o Gemini destacou-se na elaboração de cenários lúdicos e efeitos narrativos, mas foi ligeiramente inconsistente na definição clara de regras, interações e efeitos de jogo, o que comprometeu a precisão em algumas histórias estruturadas.

Tabela 3. Pontos fortes e pontos fracos identificados nas LLMs avaliadas.

Modelo	Pontos Fortes	Pontos Fracos
Llama	- Produziu respostas coerentes com o Pac-Man clássico .	- Pouca criatividade narrativa ou variação.
	- Manteve uma boa consistência estrutural nos templates.	- Cenários e interações genéricos ou previsíveis .
	- Boa precisão em respeitar as categorias Espaço, Ator, Item, etc.	- Baixa adaptação ao tom infantil e lúdico que o projeto pede.
ChatGPT	- Narrativa criativa e contextualizada , adaptando elementos de forma inventiva (ex: “Pac-Guto”, “Moeduca”).	- Às vezes priorizou a narrativa sobre a estrutura formal do template .
	- Excelente na produção de histórias coesas e personagens novos .	- Algumas descrições ficaram mais próximas de storytelling literário do que de requisitos objetivos de game story.
	- Respeitou bem a estrutura do template, adaptando-a à proposta infantil.	
DeepSeek	- Consistência técnica muito boa, seguindo a estrutura formal à risca.	- Baixa criatividade estética e narrativa.
	- Adicionou boas variações (como “chave dourada” e “fruta rara”).	- Cenários e personagens pouco inventivos .
	- Clareza na definição de interações e regras de sistema .	- Tendência a repetir soluções padrões de jogos.
Gemini	- Altíssima criatividade em cenários (“rios de bolhas”, “céu de estrelas”).	- Em alguns casos, afastou-se das características originais de Pac-Man .
	- Capacidade de criar interações e desafios lúdicos e originais.	- Um pouco inconsistente no detalhamento de regras e efeitos (ex: “usar Telescópio Mágico” sem efeito claramente definido).
	- Boa adaptação ao tom infantil, com recursos de fantasia.	

5. Conclusões e Trabalhos Futuros

Este artigo apresentou uma pesquisa voltada à análise da performance de diferentes LLMs na geração de game stories voltadas para um contexto específico, neste caso uma aventura infantil inspirada no jogo Pac-Man. O estudo teve como objetivo comparar as respostas de quatro modelos distintos — ChatGPT, DeepSeek, Gemini e Llama — utilizando um conjunto comum de prompts, de modo a avaliar as histórias geradas a partir de critérios como criatividade, jogabilidade, correteude, utilidade, pontos fortes e limitações encontradas.

Os resultados evidenciaram diferenças significativas no desempenho e nas características narrativas produzidas por cada LLM, destacando aspectos como: detalhamento e originalidade das histórias, clareza e coerência estrutural, capacidade de adaptação ao estilo infantil proposto, e potencial de aplicação prática no desenvolvimento de elementos para o design de jogos digitais. Ou seja, tratam-se de resultados que possuem implicações práticas relevantes para desenvolvedores que buscam utilizar IAs generativas em seus respectivos projetos de desenvolvimento.

Assim, no que diz respeito à abordagem e criatividade, ChatGPT e Gemini se destacaram ao apresentar universos narrativos originais, com temáticas encantadas e interações que fogem do tradicional sem comprometer a coerência e jogabilidade. DeepSeek manteve uma proposta mais próxima do clássico Pac-Man, com pequenas variações criativas, enquanto Llama optou por preservar a estrutura original sem grandes inovações. Essa diferença ficou clara nas notas atribuídas para originalidade e consistência temática, onde Gemini e ChatGPT alcançaram as pontuações mais altas. Em termos de variedade de desafios, integração entre elementos e estrutura narrativa, novamente ChatGPT e Gemini apresentaram superioridade. Ambos criaram interações claras e plausíveis, com desafios viáveis e objetivos bem definidos que guiam o jogador por um caminho de início, meio e fim. DeepSeek demonstrou competência técnica, com boa coerência lógica entre eventos e objetivos, enquanto Llama se manteve no básico, atendendo aos requisitos mínimos, mas sem explorar novas possibilidades. As médias finais atribuídas, que variaram de 4.3 (Llama) a 6.9 (ChatGPT), evidenciam essa diferença.

Como possíveis desdobramentos deste trabalho, tem-se a possibilidade de expandir as análises para incluir outros aspectos relacionados a jogos em si, tais como: balanceamento de dificuldade, feedback ao jogador e imersão narrativa. Além disso, seria pertinente desenvolver protótipos funcionais com as estruturas narrativas e interativas propostas por cada LLM (algo inclusive sugerido pelo ChatGPT ao final de sua resposta), bem como avaliar a funcionalidade, receptividade e experiência do jogador em testes controlados nos possíveis jogos gerados. Dessa forma, seria possível não apenas validar as propostas narrativas, mas também explorar o potencial de cada IA na geração procedural de aventuras personalizadas para jogos digitais educativos, infantis e de entretenimento. Por fim, a junção das diferentes respostas obtidas de modo a explorar e confirmar os ganhos de elaboração e originalidade na geração de game stories a partir da colaboração entre múltiplos LLMs [Zhao et al. 2025] também pode ser realizada em um futuro próximo.

Como outros possíveis trabalhos futuros, sugere-se aplicar o modelo de prompt estruturado em diferentes gêneros e temáticas de jogos digitais, como jogos de estratégia, terror ou aventura futurista, a fim de verificar a adaptabilidade e a consistência narrativa dos LLMs em contextos variados. Pretende-se, em estudos posteriores, incluir avaliações qualitativas e quantitativas realizadas por jogadores e por especialistas em design de jogos, a fim de validar a aplicabilidade prática e o apelo dos jogos concebidos a partir das game stories geradas. Por fim, recomenda-se conduzir uma validação estatística e/ou formal dos critérios de avaliação empregados na análise das game stories, de modo a conferir maior rigor metodológico a futuras avaliações derivadas deste trabalho.

Referências

- Alavi, S. H., Xu, W., Jojic, N., Kennett, D., Ng, R. T., Rao, S., Zhang, H., Dolan, B., e Schwartz, V. (2024). Game plot design with an llm-powered assistant: An empirical study with game designers. *arXiv preprint arXiv:2411.02714*.
- Heston, T. F. e Khun, C. (2023). Prompt engineering in medical education. *International Medical Education*, 2(3):198–205.

- Huang, L. e Sun, X. (2023). Create ice cream: Real-time creative element synthesis framework based on gpt3. 0. In *2023 IEEE Conference on Games (CoG)*, pages 1–4. IEEE.
- Junior, W. G., Marasco, E., Kim, B., Behjat, L., e Eggermont, M. (2023). How chatgpt can inspire and improve serious board game design. *International Journal of Serious Games*, 10(4):33–54.
- Lanzi, P. L. e Loiacono, D. (2023). Chatgpt and other large language models as evolutionary engines for online interactive collaborative game design. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1383–1390.
- Lund, A. M. (2001). Measuring usability with the use questionnaire12. *Usability interface*, 8(2):3–6.
- Marvin, G., Hellen, N., Jjingo, D., e Nakatumba-Nabende, J. (2023). Prompt engineering in large language models. In *International conference on data intelligence and cognitive informatics*, pages 387–402. Springer.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.
- Saito, K., Kobayashi, K., Takekoshi, W., Hashimoto, A., Hirai, N., Kimura, A., Takahashi, A., Yoshioka, N., e Mano, A. (2023). Double impact: Children’s serious rpg generation/play with a large language model for their deeper engagement in social issues. In *Joint International Conference on Serious Games*, pages 274–289. Springer.
- Sarinho, V. T. (2020). Applying user stories as game elements and interactions in a game of games design proposal. In *Proceedings of Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, pages 40–46. SBC.
- Sweetser, P. (2024). Large language models and video games: A preliminary scoping review. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, pages 1–8.
- Volum, R., Rao, S., Xu, M., DesGarennes, G., Brockett, C., Van Durme, B., Deng, O., Malhotra, A., e Dolan, W. B. (2022). Craft an iron sword: Dynamically generating interactive game characters by prompting large language models tuned on code. In *Proceedings of the 3rd Wordplay: When Language Meets Games Workshop (Wordplay 2022)*, pages 25–43.
- Zhao, Y., Zhang, R., Li, W., e Li, L. (2025). Assessing and understanding creativity in large language models. *Machine Intelligence Research*, 22(3):417–436.