

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

FELLIPE MOREIRA ROLIM

LLMs e IA Generativa em Jogos

RIO DE JANEIRO
2023

FELLIPE MOREIRA ROLIM

LLMs e IA Generativa em Jogos

Trabalho de conclusão de curso de graduação
apresentado ao Instituto de Computação da
Universidade Federal do Rio de Janeiro como
parte dos requisitos para obtenção do grau de
Bacharel em Ciência da Computação.

Orientador: Prof. Geraldo Bonorino Xexéo

RIO DE JANEIRO

2023

CIP - Catalogação na Publicação

M3191 Moreira Rolim, Fellipe
 LLMs e IA Generativa em Jogos / Fellipe Moreira
 Rolim. -- Rio de Janeiro, 2023.
 66 f.

 Orientador: Geraldo Bonorino Xexéo.
 Trabalho de conclusão de curso (graduação) -
 Universidade Federal do Rio de Janeiro, Instituto
 de Computação, Bacharel em Ciência da Computação,
 2023.

 1. Videogames. 2. LLM. 3. IA Generativa. 4. RPG.
 5. Jogos. I. Bonorino Xexéo, Geraldo, orient. II.
 Título.

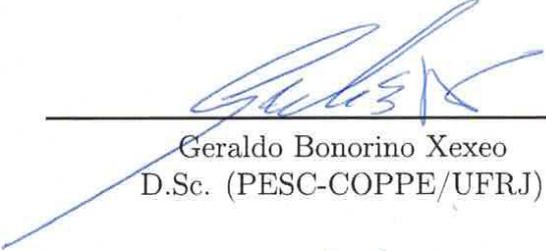
FELLIPE MOREIRA ROLIM

LLMs e IA Generativa em Jogos

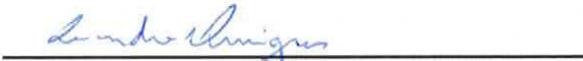
Trabalho de conclusão de curso de graduação
apresentado ao Instituto de Computação da
Universidade Federal do Rio de Janeiro como
parte dos requisitos para obtenção do grau de
Bacharel em Ciência da Computação.

Aprovado em 19 de dezembro de 2023

BANCA EXAMINADORA:


Geraldo Bonorino Xexeo
D.Sc. (PESC-COPPE/UFRJ)


Adriana Santarosa Vivacqua
D.Sc. (IC/UFRJ)


Leandro Ouriques Mendes de Carvalho
M.Sc (Marinha do Brasil)

RESUMO

Este trabalho explora a interseção entre a Inteligência Artificial (IA) e videogames, com foco especial em IA Generativa e Modelos de Linguagem Grandes (LLMs). O estudo abrange a evolução histórica da IA em jogos, uma análise aprofundada de Chatbots de IA (como o GPT-3) e LLMs, e a integração de Aprendizado de Máquina e IA Generativa nos jogos, destacando sua capacidade de se adaptar aos jogadores e proporcionar experiências personalizadas.

A incursão na IA Generativa enfatiza seu papel inovador na criação de narrativas dinâmicas e experiências de jogo únicas. Esta pesquisa explora suas aplicações práticas, desde a personalização de diálogos até a geração de ambientes dinâmicos. A implementação prática de um "Simulador de TTRPG" demonstra como essas teorias podem ser aplicadas para criar uma experiência de jogo interativa e envolvente.

As conclusões destacaram a transformação da IA de uma ferramenta reativa para uma proativa, moldando experiências mais ricas e desafiadoras. O estudo focou não apenas no estado atual da IA em videogames, mas também oferece um vislumbre de seus possíveis desenvolvimentos futuros. Considerações éticas também foram ressaltadas, sublinhando a importância de equilibrar essa inovação com preocupações sociais.

Em síntese, o relacionamento entre IA e videogames está em constante evolução, prometendo transformar a forma como jogamos e desafiando fronteiras criativas. Este estudo oferece uma imersão abrangente nesse campo dinâmico, incentivando uma atenção contínua às futuras possibilidades e desafios instigantes.

Palavras-chave: inteligência artificial; videogames; IA generativa; modelos de linguagem grandes; aprendizado de máquina.

ABSTRACT

This work explores the intersection between Artificial Intelligence (AI) and video games, with a special focus on Generative AI and Large Language Models (LLMs). The study encompasses the historical evolution of AI in games, an in-depth analysis of AI chatbots (such as GPT-3) and LLMs, and the integration of Machine Learning and Generative AI in games, highlighting their ability to adapt to players and provide personalized experiences. The foray into Generative AI emphasizes its innovative role in creating dynamic narratives and unique gaming experiences. This research explores practical applications, from dialogue customization to the generation of dynamic environments. The practical implementation of a Tabletop Role-Playing Game (TTRPG) Simulator demonstrates how these theories can be applied to create an interactive and engaging gaming experience.

The conclusions highlight the transformation of AI from a reactive tool to a proactive one, shaping richer and more challenging experiences. The study focused not only on the current state of AI in video games but also offered a glimpse into possible future developments. Ethical considerations were also underscored, emphasizing the importance of balancing innovation with social concerns.

In summary, the relationship between AI and video games is constantly evolving, promising to transform how we play and challenging creative boundaries. This study provides a comprehensive immersion in this dynamic field, encouraging ongoing attention to future possibilities and intriguing challenges.

Keywords: artificial intelligence; video games; generative AI; large language models; machine learning.

LISTA DE ILUSTRAÇÕES

Figura 1 – Máquina de Estados Finitos de uma catraca.	14
Figura 2 – Caminhos equivalentes entre A e B em um ambiente 2D	15
Figura 3 – Elementos de uma Árvore de Comportamento	15
Figura 4 – Arquitetura de Rede Neural do ChatGPT	25
Figura 5 – Técnicas de treinamento para modelos de linguagem	25
Figura 6 – Exemplo de uma função que se deseja replicar	26
Figura 7 – Possíveis funções resposta gerados pela Rede Neural	27
Figura 8 – Reajustar os pesos do modelo até chegar a um resultado aceitável	27
Figura 9 – Representação gráfica dos agrupamentos das palavras no texto de treinamento do ChatGPT, com base na frequência com que são encontradas em proximidade uma às outras	28
Figura 10 – Representação numérica de palavras no modelo do ChatGPT	29
Figura 11 – Diagrama de Fluxo do Jogo	55
Figura 12 – Telas Inicial do Jogo	56
Figura 13 – Tela de Criação de Personagem	56
Figura 14 – Tela de Criação do Mundo	57
Figura 15 – Tela de Criação da História	57
Figura 16 – Tela Principal do Jogo	58
Figura 17 – Tela do Dado	58

LISTA DE ABREVIATURAS E SIGLAS

IA	Inteligência Artificial
LLMs	Large Language Models – Modelos de Linguagem Grandes
TTRPG	Tabletop Role-Playing Game – Jogo de RPG de Mesa
NPC	Non-Player Character – Personagem Não-Jogável
Game AI	Inteligência Artificial em jogos
D&D	Dungeons and Dragons
A*	Algoritmo A* - (Algoritmo de Busca A*)
IA Diretora	Inteligência Artificial Diretora
RPG	Jogo de Interpretação de Papéis
IA Generativa	Inteligência Artificial Generativa
AI Director	Diretor de Inteligência Artificial
GPT	Generative Pre-trained Transformer – Transformador Pré-treinado Gerativo
ReLU	Rectified Linear Unit – Unidade Linear Retificada
RNN	Recurrent Neural Networks – Rede Neural Recorrente
LSTM	Long Short-Term Memory – Memória de Curto Prazo de Longo Prazo
ChatGPT	Chat Generative Pre-trained Transformer – Transformador Pré-treinado Gerativo para Conversação
ML	Machine Learning – Aprendizado de Máquina
GANs	Generative Adversarial Networks – Redes Generativas Adversariais
VAEs	Variational Autoencoders – Codificadores Automáticos Variacionais
PCG-ML	Procedural Content Generation via Machine Learning – Geração Procedural de Conteúdo via Aprendizado de Máquina
NFTs	Non-Fungible Token – Tokens Não Fungíveis
VR	Virtual Reality – Realidade Virtual

AGI	Artificial General Intelligence – Inteligência Artificial Geral
HP	Hit Points – Pontos de Vida
GPU	Graphics Processing Unit – Unidade de Processamento Gráfico
API	Application Programming Interface – Interface de Programação de Aplicações
React	Biblioteca JavaScript para construir interfaces de usuário
Vite	Ferramenta de construção front-end rápida e eficiente
JSX	JavaScript XML
UI	User Interface – Interface do Usuário

SUMÁRIO

1	INTRODUÇÃO	10
1.1	OBJETIVO GERAL	10
1.2	ESTRUTURA DO TRABALHO	11
2	INTELIGÊNCIA ARTIFICIAL EM VIDEOGAMES	12
2.1	UTILIZAÇÃO DA IA EM JOGOS	12
2.1.1	Máquinas de Estados Finitos (Finite State Machines)	13
2.1.2	Planejamento de Rotas (Pathfinding)	14
2.1.3	Scripts de Árvore de Comportamento (Behavior-Tree Scripts)	15
2.2	JOGOS SISTÊMICOS E NARRATIVA DINÂMICA	16
2.2.1	IA Diretora	16
2.2.2	Narrativa Interativa em Jogos	16
2.3	APRENDIZADO DE MÁQUINA EM JOGOS	17
2.4	GERAÇÃO PROCEDURAL EM JOGOS	18
2.5	IA E O FUTURO DO DESIGN DE JOGOS	19
3	FUNCIONAMENTO DE CHATBOTS DE IA E LLMS	21
3.1	TREINAMENTO DE CHATBOTS	22
3.1.1	Alinhamento de Chatbots	23
3.1.2	Eficácia dos Chatbots	23
3.2	REDES NEURAIS NOS LLMS	24
3.2.1	Relacionamentos e Matrizes	28
3.2.2	Atenção no Chatbot	30
3.3	FUNCIONAMENTO DE UM CHATBOT	31
4	IA GENERATIVA E O DESENVOLVIMENTO DE JOGOS	33
4.1	O QUE É IA GENERATIVA	33
4.1.1	Avanços na IA Generativa e o Mercado	35
4.1.2	IA Generativa em Jogos	36
4.2	DESAFIOS NO CAMPO DE IA GENERATIVA	37
4.2.1	Problemas Legais no Uso de IA Generativa	38
4.2.2	Como IA Generativa Pode Afetar os Desenvolvedores	39
5	USO DE LLMS E IA GENERATIVA EM JOGOS	41
5.1	DESAFIOS DE DESIGN	42
5.1.1	Limites Tecnológicos	43

5.2	UTILIZANDO LLMS EM JOGOS	44
5.2.1	Treinamento de LLMs	45
5.2.2	Modelos Grandes e Pequenos	46
5.3	INTELIGÊNCIA ARTIFICIAL GERAL (AGI)	47
5.4	COMPORTAMENTO EMERGENTE	48
6	SIMULADOR DE TTRPG - PARTE PRÁTICA	50
6.1	VISÃO GERAL DO PROJETO	50
6.2	DETALHES TÉCNICOS	51
6.2.1	Estrutura de Componentes do React	51
6.2.2	Solicitações de API	52
6.3	EXPLICAÇÃO DOS COMPONENTES	53
6.4	INTERAÇÃO COM MODELOS DE LINGUAGEM	54
6.5	TELAS DA APLICAÇÃO	54
6.5.1	Telas Inicial	55
6.5.2	Criação de Personagem	55
6.5.3	Criação do Mundo	55
6.5.4	Criação da História	56
6.5.5	Tela Principal do Jogo	56
6.5.6	Tela do Dado	57
6.6	CONCLUSÃO	58
7	CONCLUSÃO	60
7.1	CONSIDERAÇÕES FINAIS	60
	REFERÊNCIAS	62

1 INTRODUÇÃO

Devido à rápida evolução da tecnologia, a Inteligência Artificial (IA) tem desempenhado um papel cada vez mais proeminente em diversas áreas, destacando-se especialmente no campo dos videogames. A interseção entre IA e jogos transcende a simples automação dos elementos do jogo, abrangendo a criação de experiências mais dinâmicas, envolventes e personalizadas para os jogadores.

Este trabalho tem como propósito realizar uma profunda investigação sobre a aplicação da IA em videogames, com foco especial na "IA Generativa" e nos "Modelos de Linguagem Grandes" (LLMs), visando compreender não apenas o seu estado atual, mas também vislumbrar os caminhos futuros desse relacionamento.

1.1 OBJETIVO GERAL

O contexto contemporâneo dos videogames é caracterizado por uma demanda crescente por experiências de jogo mais ricas, criativas e envolventes. Nesse cenário, a aplicação da Inteligência Artificial (especialmente a IA Generativa e os LLMs) emerge como catalisadora para a transformação dessa indústria.

A motivação para este projeto reside na necessidade de compreender as nuances dessas tecnologias emergentes e seu impacto no desenvolvimento de jogos. A "IA Generativa" abre portas para a criação de mundos virtuais mais dinâmicos, enquanto os LLMs elevam a comunicação em jogos a novos níveis, permitindo interações mais naturais e imersivas entre os jogadores e os elementos dos jogos.

A parte central deste trabalho consiste na análise do impacto atual e futuro da Inteligência Artificial (IA) nos videogames, em particular a "IA Generativa" e os LLMs, e como eles impactam e continuarão a influenciar o futuro dos videogames. O objetivo principal é examinar em detalhes os mecanismos pelos quais essas tecnologias são integradas no desenvolvimento de jogos eletrônicos, destacando como elas modelam e influenciam a experiência dos jogadores. Este trabalho também se aprofunda nos seguintes aspectos:

- Investigar o papel da "IA Generativa" no desenvolvimento de narrativas e de conteúdo dinâmico em videogames, destacando como essa abordagem inovadora pode redefinir a criação de histórias interativas.
- Avaliar como os LLMs têm sido aplicados no aprimoramento da interação entre os jogadores e os jogos, explorando como esses modelos linguísticos avançados podem contribuir para a personalização e adaptação inteligente das interfaces de jogo.
- Examinar a criação e a implementação prática de um Simulador de TTRPG que implementa essas técnicas de narrativas e de conteúdo dinâmico, de modo a de-

monstrar como a IA pode criar experiências de jogo personalizadas, fornecendo aos jogadores a oportunidade de participar ativamente na criação de narrativas únicas.

1.2 ESTRUTURA DO TRABALHO

Este trabalho está estruturado em 7 capítulos, cada um contribuindo para uma visão mais abrangente da relação entre IA e videogames.

Capítulo 2: Inteligência Artificial em Videogames - Explora a influência histórica e atual da Inteligência Artificial em videogames, mostrando marcos significativos e as tendências atuais.

Capítulo 3: Funcionamento de Chatbots de IA e LLMs - Analisa o funcionamento de "Chatbots de IA" e "Modelos de Linguagem Grandes", que são fundamentais para entender as capacidades gerativas dessas tecnologias.

Capítulo 4: IA Generativa no Desenvolvimento de Jogos - Concentra-se nas aplicações da "IA Generativa" no contexto dos videogames, explorando como ela redefine a criação de narrativas e conteúdo dinâmico.

Capítulo 5: Uso de LLMs e IA Generativa em Jogos - Examina as aplicações e avanços notáveis na integração de LLMs e IA generativa em jogos.

Capítulo 6: Simulador de TTRPG - Parte Prática - Apresenta a descrição da parte prática do trabalho, o "Simulador de TTRPG", uma aplicação web que emprega "IA Generativa" e LLMs para simular um jogo de "RPG de Mesa" onde as decisões do jogador afetam dinamicamente a narrativa.

Capítulo 7: Conclusão - Conclui o trabalho consolidando as descobertas e oferecendo perspectivas sobre o papel contínuo e o futuro da inteligência artificial em videogames.

Ao longo desses capítulos, buscamos oferecer uma visão abrangente e detalhada de como a Inteligência Artificial tem evoluído, sua influência no cenário dos videogames e o seu potencial para o futuro.

2 INTELIGÊNCIA ARTIFICIAL EM VIDEOGAMES

As inteligências artificiais (IA) são utilizadas nos videogames para gerar comportamentos responsivos, adaptativos e inteligentes, principalmente em "personagens não-jogáveis"(NPCs), de modo que estes apresentem inteligência semelhante à humana.

A Inteligência Artificial tem sido parte integral dos videogames desde o seu surgimento na década de 1950. O uso de IA em videogames (Game AI) é um subcampo distinto da "IA acadêmica", com o objetivo de melhorar a experiência do jogador, em vez de se concentrar em "Aprendizado de Máquina"ou "Tomada de Decisões"(MARTINEZ, 2023).

A ideia de jogadores enfrentarem oponentes controlados por uma IA foi difundida durante a Era Dourada dos fliperamas no fim dos anos 1970 e início de 1980. Vários elementos popularizaram esta tendência, tais como níveis cuja dificuldade aumentava gradualmente, inimigos com padrões de movimento distintos e eventos dentro do jogo que dependiam das ações do jogador.

O termo "IA de Jogos"(Game AI) é utilizado para se referir ao amplo conjunto existente de algoritmos com o propósito principal de simular a inteligência humana e controlar o comportamento dos NPCs e outros elementos dos jogos. Esses algoritmos incluem diversas técnicas, como Planejamento de Rotas, Árvores de Decisão, Máquinas de Estados Finitos, etc. (YANNAKAKIS; TOGELIUS, 2018).

Deve-se notar que a Inteligência Artificial em jogos eletrônicos nem sempre pode ser considerada uma "IA verdadeira", uma vez que tais técnicas não necessariamente facilitam o aprendizado computacional ou outros critérios semelhantes, sendo muitas vezes considerada apenas uma computação automatizada.

2.1 UTILIZAÇÃO DA IA EM JOGOS

Algoritmos de IA/heurística são conjuntos de instruções ou procedimentos computacionais projetados para resolver problemas complexos por meio de abordagens baseadas em princípios de inteligência artificial e heurísticas.

Algoritmos de IA utilizam técnicas computacionais para simular o raciocínio humano, usando dados e modelos para tomar decisões e resolver problemas complexos. Heurísticas, por outro lado, empregam regras práticas baseadas em intuições para encontrar soluções aproximadas em cenários desafiadores, muitas vezes não seguindo um processo rigoroso.

Esses algoritmos são usados para uma variedade de aplicações em jogos, como tomar decisões ou realizar tarefas que normalmente seriam difíceis para um programa de computador convencional. Em jogos, algoritmos de IA/heurística são frequentemente usados para controlar NPCs, tomar decisões estratégicas, adaptar-se a situações em mudança

e melhorar a jogabilidade, de modo a criar uma experiência de jogo mais desafiadora e imersiva.

Jogos e Inteligência Artificial têm estado fundamentalmente entrelaçados desde o início, impulsionando um ao outro em direção a novas fronteiras. No seu livro "Playing Smart", o pesquisador em inteligência artificial Julian Togelius argumenta que IA e jogos estão entrelaçados de três maneiras distintas (TOGELIUS, 2019):

- Jogos impulsionaram a evolução da inteligência artificial, pois fornecem regras precisas e subdomínios restritos nos quais a inteligência artificial tradicionalmente se destaca.
- A inteligência artificial é o futuro dos videogames, seja pela adoção de "Sistemas Procedurais" robustos ou "Diretores de Narrativa" em jogos capazes de criar histórias dinâmicas instantaneamente.
- Jogos e a inteligência artificial podem nos fornecer uma compreensão mais profunda de algo ainda mais elusivo que eles mesmos, sendo isto o conceito de "inteligência" em si.

Nos primeiros jogos, a IA era bem simples, uma vez que sua IA consistia em uma combinação de "Padrões Estáticos" e "Interrupções Dinâmicas" baseadas nas ações do jogador.

Um exemplo é o jogo "Space Invaders" (Taito Corporation, 1978), onde a IA é integrada a um design no qual a limitação no hardware do jogo fazia com que os inimigos ficassem mais rápidos à medida que eram derrotados, criando uma sensação dinâmica de progresso conforme o jogador progredia no jogo (OSBOURN, 2023).

2.1.1 Máquinas de Estados Finitos (Finite State Machines)

O método mais comum utilizado para Inteligência Artificial em jogos é o uso de "Máquinas de Estados Finitos" (Finite State Machines), no qual se organiza um jogo em uma série de estados, cada qual possui um conjunto de instruções específicas. A IA é então direcionada para um desses estados com base nas ações do jogador ou pela ativação de um "gatilho" pré-estabelecido (HU; ZHANG; MAO, 2011).

Um exemplo disso é no jogo "Pac-Man" (Namco, 1980); no estado neutro, cada um dos fantasmas segue scripts diferentes sobre como perseguir o Pac-Man. Um deles o persegue normalmente, outro foca-se nos dois quadrados à frente do Pac-Man. Outro combina esses dois métodos. E finalmente, o fantasma laranja recua se chegar muito perto do Pac-Man. Quando o Pac-Man come uma "pílula de poder", todos os inimigos começam a fugir dele, temos então que os inimigos entram em um novo estado com novos comportamentos. O sistema de IA simplesmente alterna entre esses modos com base no estado atualmente

acionado, e isso pode ser conceituado de maneira bastante simples. Com esse conjunto simples de comportamentos, o Pac-Man gerava uma jogabilidade envolvente e profunda que se ajustava dinamicamente às ações dos jogadores.

Um exemplo de uma Máquinas de Estados Finitos, os estados de uma catraca: Inserir uma moeda (Coin) em uma catraca a destranca, e após a catraca ser empurrada (Push), ela trava novamente. Inserir uma moeda em uma catraca destrancada ou empurrar contra uma catraca travada não alterará seu estado. A Figura 1 abaixo ilustra os estados e uma catraca e as transições entre estes estados.

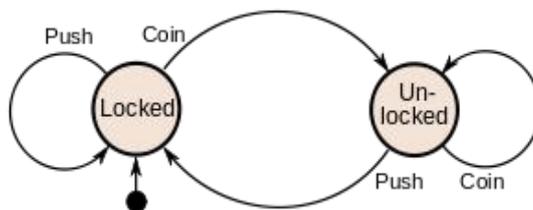


Figura 1 – Máquina de Estados Finitos de uma catraca.

2.1.2 Planejamento de Rotas (Pathfinding)

O objetivo do algoritmo de Planejamento de Rotas (Pathfinding) é encontrar o caminho mais eficiente através dessa grade, evitando possíveis obstáculos ao longo desse caminho (GAMES, 2014). Uma variante desse algoritmo é o Planejamento de Rotas em Grade (Grid Pathfinding) que envolve dividir o espaço do jogo em uma grade de quadrados uniformes.

Os jogos stealth (furtivos) são casos que normalmente empregam inimigos que seguem scripts de “Planejamento de Rotas” (Pathfinding) até que suas ações ou presença os forcem a um “estado de confronto”, sendo que o “Planejamento de Rotas” em jogos normalmente é tratado por um “Algoritmo A*”.

O “Algoritmo A*” é um método que pode ser empregado nesse algoritmo, o qual avalia os quadrados de uma grade com base em duas métricas: o “custo estimado” do caminho até o destino (chamado de heurística) e o “custo real” do caminho percorrido até a célula atual. O algoritmo A* explora os quadrados da grade de acordo com essas métricas e escolhe o caminho mais promissor, eventualmente encontrando o caminho mais curto.

A maioria dos jogos usa variantes de “Máquina de Estados Finitos” e “Algoritmos A*” para configurar os elementos da sua IA, e muitas vezes isso é suficiente para criar uma jogabilidade cativante.

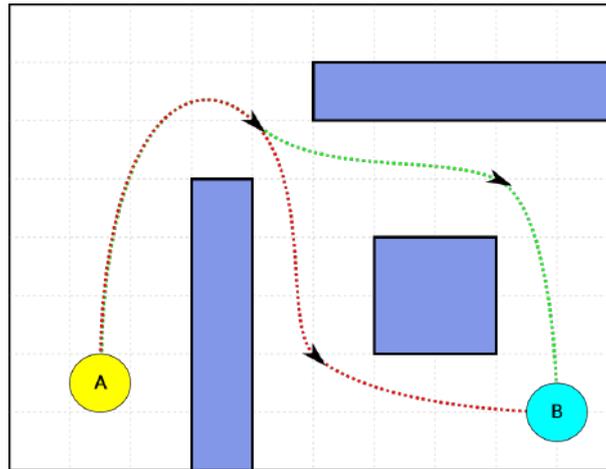


Figura 2 – Caminhos equivalentes entre A e B em um ambiente 2D

2.1.3 Scripts de Árvore de Comportamento (Behavior-Tree Scripts)

Uma variação bastante usada da “Máquina de Estados Finitos” são os “Scripts de Árvore de Comportamento” (Behavior-Tree Scripts). Estas árvores são bastante semelhantes às “Máquinas de Estados Finitos”, exceto que elas são hierárquicas, criando uma ordem de prioridade para os comportamentos exibidos pela IA (SIMPSON, 2014).

Uma Árvore de Comportamento é uma estrutura de dados hierárquica composta por "nós" que geralmente possuem vários “nós-filhos”. Cada nó especifica uma sequência de ações a serem tomadas ou pode aplicar operações lógicas (como condicionais e negações), sendo a prioridade para qual ramo será selecionado é calculada com base no que está acontecendo no jogo.

Essa técnica foi popularizada pelo estúdio Bungie com o jogo "Halo 2" (Microsoft Game Studios, 2004), sendo o seu propósito original o de lidar com o crescente problema da complexidade na criação de inimigos (ISLA, 2005).

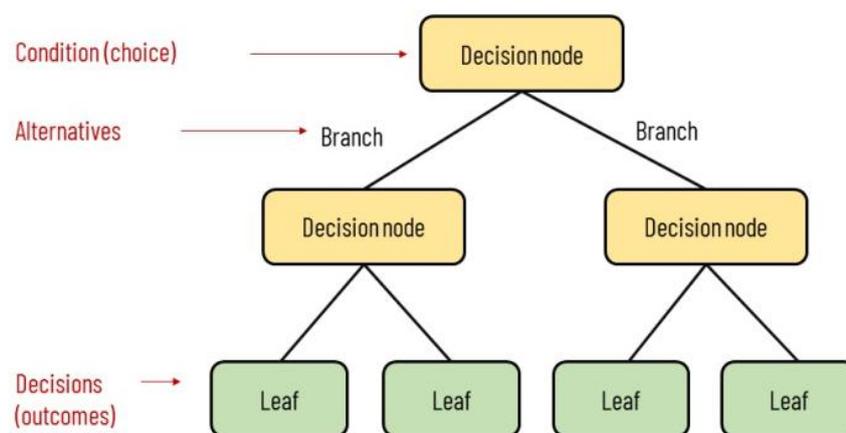


Figura 3 – Elementos de uma Árvore de Comportamento

2.2 JOGOS SISTÊMICOS E NARRATIVA DINÂMICA

Os “Jogos Sistêmicos” são jogos projetados para serem emergentes, ou seja, para surgirem das interações complexas entre múltiplos sistemas simples. Nesses jogos, a IA é frequentemente usada para gerar comportamentos emergentes, os quais podem ser previsíveis e imprevisíveis. Com “Jogos Sistêmicos”, começamos a tentar entender o funcionamento da IA a partir de um nível mais alto de abstração.

2.2.1 IA Diretora

No jogo "Left 4 Dead" (Valve, 2008), o jogador luta contra legiões de mortos-vivos em cooperação com outros jogadores. O mesmo emprega uma “IA Diretora” (AI Director) que monitora a tensão sentida pelo jogador e ajusta a experiência do jogo de acordo. Assim, o jogo utiliza uma IA para gerar comportamentos emergentes (BOOTH, 2009).

Agora, então, estamos pensando na IA no contexto de algo semelhante a um autor ou contador de histórias. A “IA Diretora” desse jogo monitora a habilidade do jogador e o número de inimigos que ele está enfrentando, e então determina quando enviar mais inimigos e quais tipos. Nota-se que agora temos um sistema rudimentar para tentar “ler” o jogador e gerenciar a experiência do mesmo, não apenas programar o comportamento da IA.

A ideia de que podemos aproveitar a IA para gerar uma experiência narrativa única é algo que dificilmente poderia ter sido previsto no início dos jogos, mas aqui vemos nossas primeiras incursões nesse domínio.

2.2.2 Narrativa Interativa em Jogos

A criação de “regras interativas” que eventualmente permitam aos usuários interagirem com ambientes virtuais de forma mais natural e realista enfrenta vários problemas, que normalmente estão relacionados a limitações dos tipos de entradas, reconhecimento e gerenciamento de sistemas.

Em sua palestra “O Futuro da Narrativa”, Jesse Schell diz que jogos precisam começar a empregar técnicas de “reconhecimento de texto para voz” e “reconhecimento de fala”, se desejam viés mais robustas para a narrativa. Da mesma forma, também é importante estudarmos a estrutura da “narrativa modular”, de modo entender como efetivamente gerenciar “IA Diretoras” e seus sistemas. No momento, ainda estamos utilizando sistemas estáticos para contar histórias em jogos (SCHELL, 2013).

Um exemplo da aplicação de uma “narrativa dinâmica” pode ser vista no jogo "Façade" (Procedural Arts, 2005), onde o jogador assume o papel de um amigo de um casal em conflito. A “IA Diretora” constrói a narrativa do jogo, escolhendo de forma inteligente o próximo ponto da história com base nas interações do jogador e na necessidade de satisfazer a história em si (EL-NASR, 2013).

O jogo também possui um “analisador de texto” (text parser) que permite ao sistema reconhecer as entradas do jogador e um “planejador reativo” (reactive planner) que controla o comportamento momento-a-momento dos personagens, misturando dinamicamente vários comportamentos simultâneos. Embora este seja um sistema bastante rudimentar, ele nos apresenta um modelo do que pode ser feito em maior escala no futuro.

2.3 APRENDIZADO DE MÁQUINA EM JOGOS

O “Aprendizado de Máquina” (Machine Learning) é um subcampo da Inteligência Artificial que se concentra no desenvolvimento de algoritmos e modelos capacitando sistemas computacionais a aprender padrões e tomar decisões, sem serem explicitamente programados para tarefas específicas. Em vez de seguir instruções programadas, os sistemas de aprendizado de máquina utilizam dados para aprender e melhorar sua performance ao longo do tempo.

No livro "The Master Algorithm", Pedro Domingos argumenta que o futuro da Inteligência Artificial está intimamente ligado à capacidade de aprendizado das IAs, uma vez que elas são capazes de adquirir conhecimento a partir de dados e aprimorar seu desempenho progressivamente (AUERBACH, 2015).

Ele apresenta a ideia de que não podemos depender apenas dos sistemas de IA "estáticos" criados até agora e que devemos capacitar os sistemas de IA a aprenderem por si mesmos, a fim de avançarmos em direção a novos horizontes. De acordo com Domingos, o "Aprendizado de Máquina" existe em cinco tipos, cada um replicando um processo natural do mundo real:

1. **Raciocínio Indutivo:** essa abordagem envolve extrapolar conclusões gerais a partir de exemplos específicos. É uma forma de aprendizado baseada em observações e inferências.
2. **Método Científico:** nessa abordagem, modelos são criados e testados por meio de experimentação e observação. Os modelos são ajustados para se adequar aos dados, e as previsões são feitas a partir desses modelos ajustados.
3. **Conexismo:** se inspira na estrutura do cérebro humano, modelando a aprendizagem como uma rede neural. Ela envolve a criação de redes de neurônios artificiais que podem aprender e se adaptar a partir de dados.
4. **Algoritmos Evolutivos:** essa abordagem é inspirada pela evolução biológica e usa algoritmos para simular a seleção natural. Os melhores "indivíduos" em uma população de soluções candidatas são selecionados e combinados para gerar novas soluções.

5. **Teorema de Bayes:** se baseia na teoria das probabilidades e na atualização de crenças. Ela envolve a atualização de probabilidades a partir de novos dados, a fim de obter uma estimativa mais precisa de eventos futuros.

Em seu livro, Togelius diz que, se desejamos realizar a visão de jogos impulsionados por IA, precisaremos de uma IA capaz de adaptar seu comportamento, aprender, compreender e até mesmo criar novos jogos. Ele introduz como o "Aprendizado de Máquina" pode ser aplicado em jogos, abordando também o tema de "Redes Neurais" e "Aprendizes por Reforço", explicando como eles empregam o processo de "retro-propagação" para treinar a IA a tomar decisões apropriadas. As "redes neurais", por exemplo, podem ser aplicadas para ensinar a IA a reconhecer padrões como rostos e personagens, além de auxiliar na composição musical (RISI; TOGELIUS, 2017).

O "Aprendizado de Máquina" é atualmente a nova fronteira da pesquisa em Inteligência Artificial. É interessante pensar nas maneiras em que ele pode melhorar os jogos e ser estudado por meio de jogos. No entanto, é vital reconhecer que os jogos são construídos com base em regras e sistemas geradores consistentes, tornando a compreensão da estrutura sistêmica dos jogos um elemento essencial para o avanço nessa área.

2.4 GERAÇÃO PROCEDURAL EM JOGOS

Ao analisarmos jogos onde os sistemas de narrativa são gerados proceduralmente, como "Dwarf Fortress" (Bay 12 Games, 2006) e "RimWorld" (Ludeon Studios, 2018), obtemos uma melhor compreensão de como a narrativa dinâmica e sistemas convergem. Estamos gerando regras que, por sua vez, geram histórias. O jogo "RimWorld" oferece ao jogador opções de diferentes "diretores de narrativa" que alteram como os eventos sistêmicos do jogo podem ocorrer. Isso demonstra o poder das ferramentas de IA para gerenciar diferentes possibilidades de jogabilidade e narrativa (HENDRIKX; MEIJER; VELDEN, 2013).

Atualmente, "ferramentas procedurais" são comumente utilizadas em vários jogos, como "Minecraft" (Mojang, 2011) e "Spelunky" (Mossmouth, 2008). Elas são utilizadas para gerar níveis de forma aleatória e proporcionar uma experiência de jogo dinâmica e inovadora. Esse método de desenvolvimento de jogos tem raízes no jogo "Rogue" (Michael Toy, Glenn Wichman, Ken Arnold, 1980), que foi pioneiro na automatização da criatividade em jogos. Inspirados pela narrativa dinâmica de Dungeons and Dragons (DD), os designers de "Rogue" optaram por empregar ferramentas procedurais para replicar a sensação de novidade e aventura que experimentavam em seus jogos de D&D.

O algoritmo em "Rogue" opera da seguinte maneira: primeiro ele cria uma "dungeon", dividindo-a em segmentos, criando salas e marcando aquelas que são visitadas pelo jogador. Em seguida, o algoritmo cria conexões entre essas salas para permitir a travessia.

Posteriormente, são adicionados itens e monstros para criar a jogabilidade neste mundo procedural (TOGELIUS, 2011).

Outro jogo que emprega ferramentas procedurais para criar um universo virtual vasto e expansivo é o jogo "Elite"(David Braben e Ian Bell, 1984). Sua utilização de métodos de geração procedural influenciou jogos como "Minecraft" e "No Man's Sky"(Hello Games, 2016). Em "Elite", uma técnica conhecida como "Valor de Semente"(seed value) é utilizada para gerar dinamicamente diferentes áreas do universo, sem a necessidade de armazenar fisicamente milhares de sistemas estelares.

Essa "semente"(seed) atua como um código que codifica o espaço emergente e, a partir de um conjunto simples de diretrizes e códigos, constrói ao seu redor um novo espaço virtual instantaneamente. Isso permite que os jogos explorem vastos universos de maneira eficiente, sem sobrecarregar o armazenamento de dados, e oferecem uma experiência de jogo variada para cada nova iteração, devido à natureza procedural da criação do mundo virtual (HENDRIKX; MEIJER; VELDEN, 2013).

Com a geração procedural, obtemos uma visão de como a própria Inteligência Artificial pode ser usada para criar jogos, o que levanta a questão de como fazer com que a IA demonstre "verdadeira" criatividade.

Essas realizações sobre IA generativa levantam uma questão importante sobre a autoria: em que ponto a IA e os seres humanos se diferenciam?

Já existem programas capazes de enganar as pessoas, fazendo-as acreditar que artes e músicas criadas por computador são autenticamente humanas. Portanto, é concebível que, em um futuro próximo, a própria IA possa ser reconhecida como um autor e ser indicada para prêmios em design de jogos, reforçando ainda mais a fronteira entre a criatividade humana e a criatividade artificial.

2.5 IA E O FUTURO DO DESIGN DE JOGOS

Para alcançarmos um futuro impulsionado por IA, destaca-se a necessidade de personagens que possam reconhecer as ações dos jogadores e responder a elas de forma significativa. Esses personagens também devem possuir defeitos e vulnerabilidades que afetem suas interações dentro do jogo, bem como a capacidade deles de fazer escolhas que tenham impacto na narrativa de forma dinâmica.

Atualmente, tornou-se uma prática comum empregar "adereços reativos" em jogos, substituindo, assim, personagens genuínos. Essa mudança não se limita apenas à aplicação de inteligência artificial em contextos de combate, mas se estende à criação de universos simulados dinâmicos (TOGELIUS, 2019).

Isso nos leva à questão de como seria o futuro dos jogos se adotássemos "verbos não-violentos"(ações ou interações que os jogadores podem realizar em um jogo que não envolvem violência ou conflito direto) como prática comum. Isso poderia resultar em jogos

nos quais as interações com o ambiente e os personagens seriam mais ricas e significativas, permitindo um amplo conjunto de escolhas e resultados. Essa abordagem poderia levar a narrativas mais dinâmicas e imersivas, onde os jogadores desempenham um papel ativo na criação da história do jogo. Isso seria uma mudança inovadora, que nos abriria novas possibilidades criativas e expandiria a forma como os jogos são vistos e apreciados.

A Inteligência Artificial é o futuro dos jogos, mas ao mesmo tempo, jogos também representam o futuro da inteligência artificial. Ambos podem oferecer "insights" sobre a inteligência em si, e além disso, podem nos proporcionar uma reflexão sobre nossa própria humanidade.

A relação entre a IA e os jogos é uma história de colaboração mútua, inspiração interligada e sinergia simbiótica. Não há razão para acreditar que o futuro será diferente. Ambos os campos continuarão a se influenciar mutuamente, resultando em avanços emocionantes e inovações que moldarão o cenário dos jogos e da inteligência artificial.

3 FUNCIONAMENTO DE CHATBOTS DE IA E LLMS

Um “Chatbot de IA” é um programa que utiliza Inteligência Artificial (como um Modelo de Linguagem Grande) para interagir e conversar com usuários de forma natural, simulando uma conversa humana. Eles normalmente são treinados em grandes conjuntos de dados textuais para serem capazes de entender perguntas, fornecer respostas úteis e realizar tarefas, tudo por meio de linguagem natural (ADAMOPOULOU, 2020).

O ChatGPT é um exemplo de “Chatbot de IA” desenvolvido pela OpenAI, sendo atualmente o aplicativo cujo consumo teve o crescimento mais rápido na história. No entanto, compreender o funcionamento do ChatGPT e de outros chatbots pode ser uma tarefa bastante desafiadora. Exploraremos então o que sabemos sobre como o ChatGPT (e Chatbots de IA em si) funciona. Isso inclui informações sobre como ele foi criado, seu processo de funcionamento real e também o futuro da sua tecnologia.

O ChatGPT é um “Chatbot de IA”, que é um “Modelo de Linguagem Grande” (Large Language Model - LLM), produzido pela OpenAI, uma organização de pesquisa em inteligência artificial que procura criar e avançar as tecnologias nessa área (LOCK, 2022).

A sigla “GPT” significa “Generative Pre-trained Transformer” (Transformador Pré-treinado Generativo):

- **Generativo:** o chatbot é capaz de gerar texto. Dado um “prompt”, ele pode produzir ou continuar um texto de maneira coerente e contextualmente relevante.
- **Pré-treinado:** ele é pré-treinado em um grande corpus de dados de texto, antes de ser ajustado para tarefas específicas.
- **Transformador:** o “Transformer” se refere à arquitetura de “Rede Neural” que ele emprega, projetada para processamento eficiente de sequências em tarefas de linguagem, usando mecanismos de auto-atenção para capturar os relacionamentos e dependências entre palavras.

Combinando esses componentes acima, temos o “GPT”, um modelo de linguagem capaz de gerar textos semelhantes aos feitos por humanos, usando seu conhecimento pré-treinado de padrões de linguagem e a sua arquitetura “Transformer” para capturar nuances de contexto e semântica dentro desses textos (VASWANI, 2017).

O GPT tem alcançado resultados impressionantes em várias tarefas de “processamento de linguagem natural” (tarefas que envolvem o uso de IA para entender, analisar e gerar linguagens humanas) e se tornou um dos pilares do processamento de linguagem de inteligência artificial moderno.

A função básica de qualquer “Modelo de Linguagem Grande” (LLM) é a de realizar um treinamento extensivo em uma grande quantidade de texto e, posteriormente, ge-

rar saídas (dado algum input) que se assemelhem ao texto utilizado no seu treinamento (BRUNDAGE, 2023).

O treinamento extensivo permite que esses modelos compreendam padrões complexos, relações entre as palavras e até mesmo a sutileza das intenções por trás do texto. Após a conclusão do treinamento, os LLMs, como o ChatGPT, tornam-se capazes de gerar saídas coerentes e relevantes quando apresentados com um input específico. Em outras palavras, eles têm a habilidade de produzir respostas que se assemelham ao estilo e ao conteúdo do texto no qual foram treinados.

Essa capacidade de geração de texto é importante nos cenários de interação humano-máquina, tais como em chatbots, onde os usuários podem se comunicar de forma natural com a IA. Os LLMs, ao compreenderem a estrutura da linguagem, podem oferecer respostas contextualmente relevantes, tornando a interação mais fluida e intuitiva para os usuários. Essa adaptabilidade e capacidade de produção de texto coerente são características distintivas que destacam a eficácia dos “Modelos de Linguagem Grandes” na criação de “Chatbots de IA” avançadas.

3.1 TREINAMENTO DE CHATBOTS

A tecnologia dos “Chatbots de IA” é incrivelmente fascinante para muitas pessoas devido à sua habilidade de interagir com os usuários de forma semelhante a um humano. No entanto, é importante compreender que os “Modelos de Linguagem Grandes” alcançaram esse nível de sofisticação ao analisarem uma vasta quantidade de dados de texto.

De acordo com o artigo sobre a criação do GPT-3, esse modelo foi treinado em mais de 500 Gigabytes de dados de texto provenientes de diversas fontes, como internet, livros digitalizados, Wikipedia, etc. Essencialmente, esse modelo foi treinado em bilhões de páginas web, com trilhões de sentenças de texto, bem como mais de 500 milhões de livros digitalizados, contendo bilhões de palavras (BROWN et al., 2020).

Como se pode imaginar, o treinamento de um “Modelo de Linguagem” com uma quantidade tão vasta de texto consome uma quantidade significativa de tempo e recursos financeiros. O modelo funcional do ChatGPT surgiu após o mesmo ter processado trilhões de palavras, utilizando supercomputadores operando em paralelo durante meses.

Após essa etapa, esses sistemas computacionais estabeleceram 107 bilhões de “conexões” entre todas essas palavras. Todas essas conexões precisam ser recalculadas sempre que alguém faz uma pergunta input ao ChatGPT, sendo essa complexidade uma das razões pelas quais o treinamento de um “Modelo de Linguagem Grande” como o ChatGPT implica custos na ordem de bilhões de dólares. Operar esse chatbot para 100 milhões de usuários ativos por mês pode resultar em despesas da ordem de meio milhão de dólares por dia (MAGUIRE, 2023).

Temos então que um “Chatbot de IA” (como o ChatGPT) é um “Modelo de Linguagem

Grande” que passou por um treinamento extenso em uma quantidade gigantesca de texto, onde durante todo esse processo ele estabeleceu bilhões de “conexões” entre as palavras presentes nesses textos. Como resultado, ele possui a capacidade de gerar (de forma impressionantemente coerente) uma continuação razoável para qualquer texto fornecido a ele como entrada.

As respostas geradas por este modelo adquirem uma aparência mais natural e semelhante à de um humano quando se introduz uma pequena dose de ‘aleatoriedade’ na escolha da próxima palavra para continuar a entrada. Essa é, portanto, a primeira conclusão fundamental sobre os “Chatbots de IA”: seus processos essenciais envolvem a adição incremental de uma palavra de cada vez à entrada, referida como ‘prompt’.

3.1.1 Alinhamento de Chatbots

Como o Chatbot faz para garantir que as palavras estejam alinhadas com o que um ser humano consideraria razoável? Esse é um dos problemas mais notórios e ainda não resolvidos no campo de inteligência artificial: o Alinhamento (V, 2023).

O desafio do Alinhamento envolve a busca de técnicas e maneiras para fazer com que uma IA compartilhe e valorize os mesmos princípios e valores que nós, seres humanos, fazemos.

Temos então que chatbots, como o ChatGPT, não são apenas treinados em palavras, mas também no quão bem sua seleção de palavras se alinha com os valores reconhecidos como ‘úteis’. Esse alinhamento é implementado por meio de um “Aprendizado por Reforço a partir do Feedback Humano” durante o treinamento do modelo.

Para o ChatGPT, a OpenAI contratou colaboradores para avaliar as respostas dadas pelo ChatGPT e usou essas respostas para criar um segundo modelo que recompensava o ChatGPT por gerar texto alinhado com as sensibilidades humanas (basicamente, um reforço positivo). Embora o resultado não seja perfeito, isso é ao menos uma tentativa de resolver o problema que apresenta resultados.

3.1.2 Eficácia dos Chatbots

Neste ponto, temos um modelo treinado em uma quantidade de texto que ultrapassa a capacidade de leitura de qualquer ser humano, além de salvaguardas que buscam priorizar os valores humanos e uma interface de usuário simples e intuitiva.

Naturalmente, essa produção massiva não seria útil ou interessante se o conteúdo do chatbot em si não fosse de qualidade. O ChatGPT é surpreendentemente eficaz na geração de respostas que se assemelham às humanas, até mesmo para perguntas difíceis. O modelo demonstrou ter uma capacidade avançada o bastante para realizar tarefas complexas, tais como passar em exames como o de licenciamento médico nos Estados Unidos, entre outros.

Apesar desses chatbots serem considerados 'inteligentes' em termos de conhecimento, é importante ressaltar que os “Modelos de Linguagem Grandes” não podem ser considerados como se eles realmente 'soubessem' algo de maneira apropriada. Também é aconselhável destacar que eles não devem ser confiados para tarefas de extrema importância.

Além disso, é fundamental compreender que, apesar de qualquer impressão que possa ser causada, os “Chatbots de IA” e outros modelos de linguagem não possuem consciência. Quando o usuário não está fazendo uma pergunta direta ao modelo, não há nenhuma atividade mental ocorrendo em sua 'mente'. Ele é estático e desprovido de pensamento ou consciência (NEWPORT, 2023).

3.2 REDES NEURAIIS NOS LLMS

Como os chatbots decidem quais palavras usar e como compreendem o contexto quando fornecidos com um 'prompt'? Isso é possível graças à arquitetura subjacente dos chatbots e de outros “Modelos de Linguagem Grandes”, conhecida como “Rede Neural”. O termo é usado porque essa arquitetura se assemelha aos neurônios e à rede neural presente no cérebro humano.

Os cérebros humanos contêm aproximadamente 100 bilhões de neurônios, e cada um desses neurônios pode ter até cerca de 1000 conexões com outros neurônios, permitindo a transmissão de sinais elétricos entre essas conexões até 1000 vezes por segundo. A distinção crucial é que esses sinais não são aleatórios; eles dependem das conexões existentes entre os neurônios e da força dessas conexões.

As “Redes Neurais” artificiais são montadas de maneira semelhante a esses neurônios, onde os neurônios artificiais interconectados enviam sinais dependendo da força (pesos) dessas conexões. A Figura 4 abaixo ilustra a estrutura dessa arquitetura neural. (CRETU, 2023)

No caso do ChatGPT, a “Rede Neural” empregada é composta por cerca de bilhões de pesos distintos, obtidos através de um extenso processo de treinamento. Quando o modelo multiplica esses pesos entre si, eles determinam, em última instância, qual palavra o modelo terá a maior probabilidade de adicionar em seguida. (MAGUIRE, 2023)

Os cientistas determinam o valor correto para os 'pesos' do modelo de maneira simples. Eles fornecem ao modelo o máximo de exemplos possível e ajustam os pesos até que a saída produzida pelo modelo se assemelhe aos exemplos.

Input → Output = Example

O treinamento de uma “Rede Neural” é realizado principalmente por meio de dois métodos amplamente adotados para a previsão de palavras: (EDUCATION), 2023)

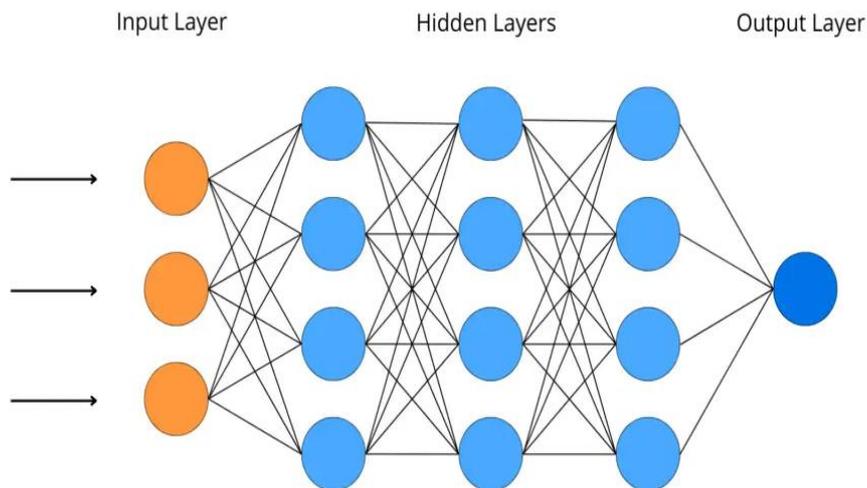


Figura 4 – Arquitetura de Rede Neural do ChatGPT

- **Previsão do Próximo Token:** é uma tarefa de “modelagem de linguagem” em que um modelo prevê a próxima palavra em uma sequência de palavras, auxiliando na geração de texto coerente e contextualmente relevante.
- **Modelo de Linguagem Mascarada:** envolve ‘mascarar’ ou ocultar certas palavras em uma sentença, e o modelo aprende a prever as palavras ausentes com base no contexto das palavras circundantes.

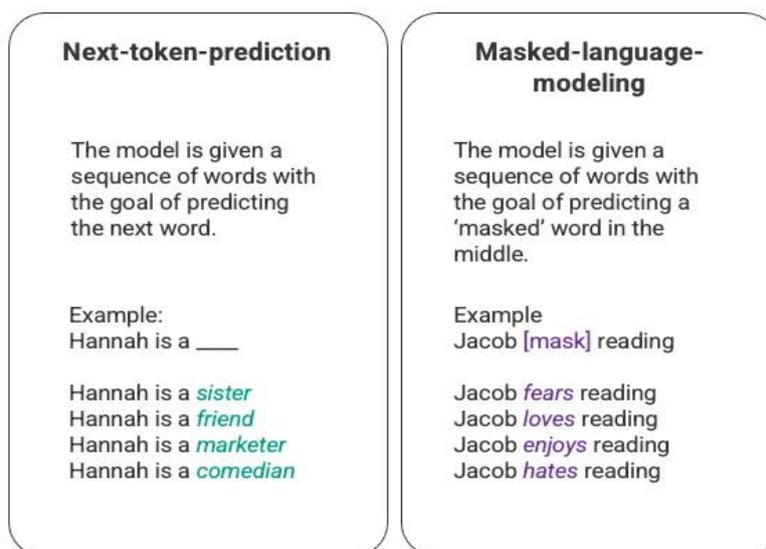


Figura 5 – Técnicas de treinamento para modelos de linguagem

A Figura 5 acima ilustra alguns exemplos de como essas técnicas de treinamento funcionam. Imagine que fornecemos uma frase incompleta a um modelo e pedimos para

ele preenchê-la. Com base nos seus pesos atuais, os quais foram ajustados por meio de um extenso conjunto de textos de treinamento, o modelo gera uma lista de probabilidades para cada palavra que ele conhece e seleciona a palavra mais provável para completar a frase.

No entanto, como pode-se avaliar a precisão ou imprecisão da resposta do modelo em um exemplo específico, considerando suas ponderações atuais? A resposta a essa pergunta é obtida por meio da aplicação de princípios matemáticos.

Podemos ilustrar esse processo imaginando que cada palavra conhecida pelo modelo é associada a um valor numérico. Se o modelo escolher uma palavra que seja significativamente diferente (em termos de valor numérico) das outras palavras que ocorrem com frequência nos dados de treinamento, técnicas estatísticas podem ser empregadas para ajustar os pesos do modelo. Isso visa tornar as escolhas de palavras mais consistentes com os dados presentes no processo de treinamento.

Em essência, o que estamos constantemente buscando é encontrar os pesos que permitam que a “Rede Neural” reproduza os exemplos que fornecemos. Estamos, portanto, confiando na capacidade da “Rede Neural” de interpolar, de maneira coerente, entre esses exemplos. Para ilustrar esse conceito, consideremos um exemplo simples, onde tentamos ensinar uma “Rede Neural” a replicar a seguinte função (Figura 6): (WOLFRAM, 2023)

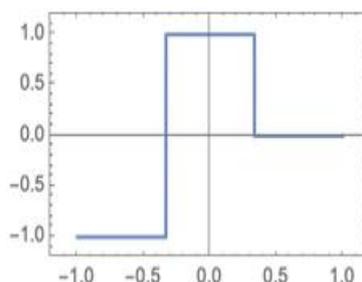


Figura 6 – Exemplo de uma função que se deseja replicar

Agora, quais pesos devem ser usados? Para cada conjunto possível de pesos, a “Rede Neural” calculará uma função possível como resposta. A cada suposição que o modelo faz, podemos calcular facilmente a diferença entre os pontos, simplesmente examinando as coordenadas X e Y da função de resposta e comparando-as às da entrada.

Podemos ver nos exemplos (Figura 7) acima que nenhum deles chega perto de reproduzir a função que queremos. Como então podemos encontrar os pesos que possam reproduzir essa função? A ideia básica é fornecer um número suficiente de exemplos “Input → Output” para permitir que o modelo aprenda e, em seguida, tentar encontrar os pesos que sejam capazes de reproduzir esses exemplos.

Depois, iteramos ajustando repetidamente os pesos do modelo, tentando reduzir o erro, ou o que é conhecido na estatística como ‘perda’ (*loss*), um processo que é repetido milhões de vezes, refinando assim os pesos da “Rede Neural”.

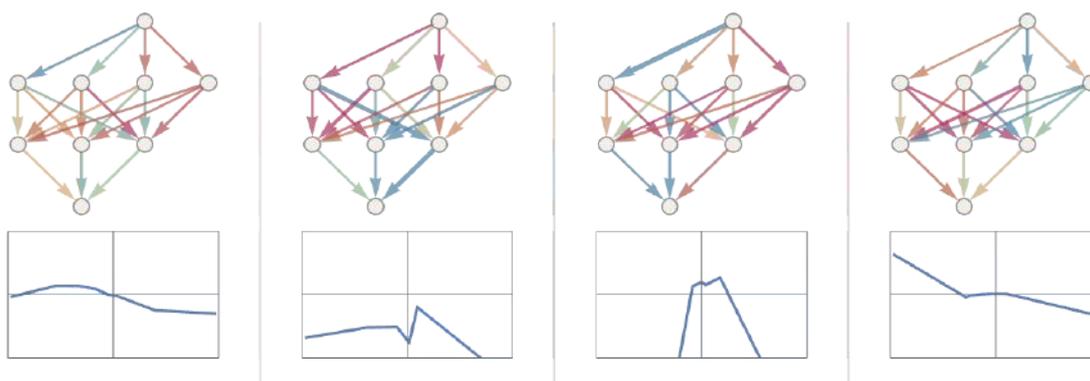


Figura 7 – Possíveis funções resposta gerados pela Rede Neural

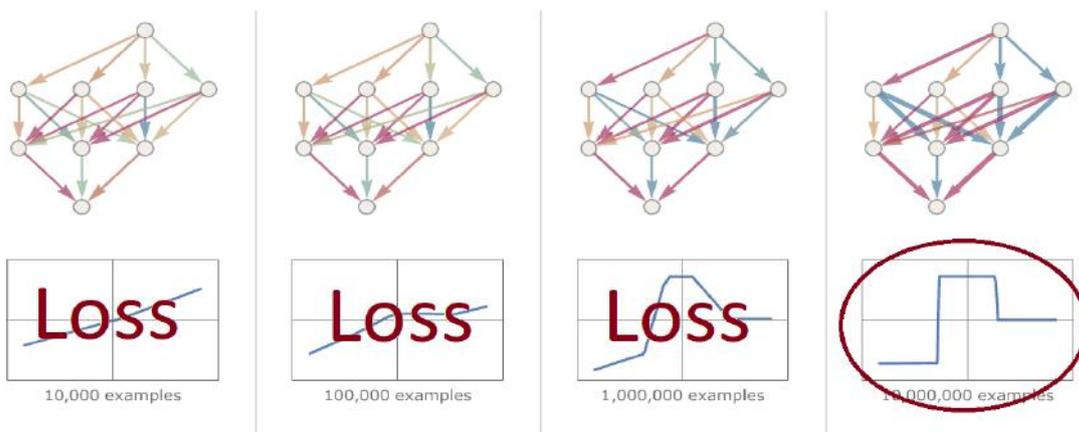


Figura 8 – Reajustar os pesos do modelo até chegar a um resultado aceitável

Se o modelo estiver funcionando, ao longo do tempo a perda será minimizada e o modelo começará a reproduzir os exemplos de forma apropriada, que neste caso específico é a função matemática correta (Figura 8).

Como pode-se aplicar esse princípio a palavras? Isso pode ser feito por um processo de tentativa e erro. O ChatGPT foi treinado atribuindo um valor numérico único para cada palavra no idioma inglês, com cerca de 50 mil palavras, cada uma com seu respectivo valor numérico correspondente. Portanto, quando fazemos uma pergunta ao modelo, ele traduz sua consulta para o equivalente em 'palavras numeradas' e, em seguida, combina esses números com os bilhões de pesos diferentes que ele aprendeu durante o seu treinamento. Isso resulta em uma lista de todas as palavras no idioma inglês, com suas probabilidades associadas, e o modelo escolhe a palavra com a maior probabilidade como resposta. (MAGUIRE, 2023)

Se fizermos isso por tempo suficiente, podemos monitorar a diferença numérica entre os exemplos fornecidos ao modelo e as respostas que ele gera, permitindo-nos minimizar essa perda ao longo do tempo. Esse método nos leva a criar algo que soa exatamente como algo que poderia ser escrito por seres humanos.

3.2.1 Relacionamentos e Matrizes

Precisamos agora de um método para associar as palavras umas às outras. Uma maneira de realizar essa tarefa é por meio de números.

Se atribuirmos um número a cada palavra na língua inglesa e, com base em análises estatísticas, determinarmos com qual frequência cada palavra ocorre ao lado de outra nos textos de treinamento fornecidos, poderíamos essencialmente criar um gráfico no qual as palavras não são distribuídas aleatoriamente; em vez disso, elas são agrupadas em conjuntos com base na frequência com que são encontradas em proximidade no texto de treinamento, como visto na Figura 9 (WOLFRAM, 2023).

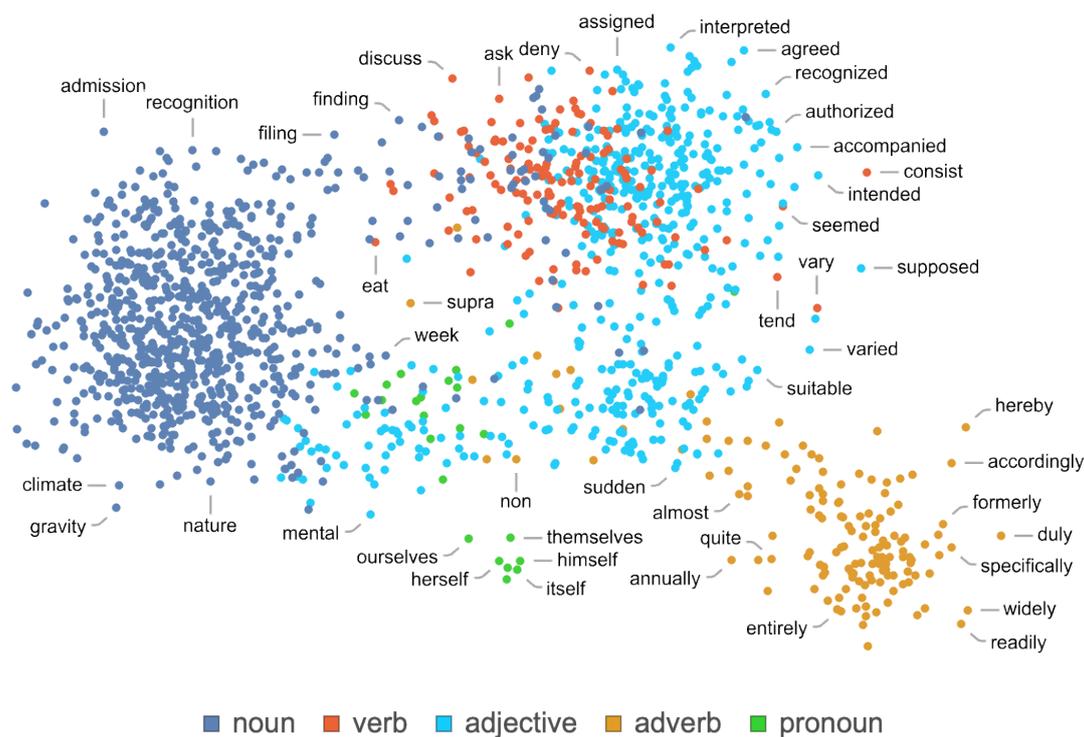


Figura 9 – Representação gráfica dos agrupamentos das palavras no texto de treinamento do ChatGPT, com base na frequência com que são encontradas em proximidade uma às outras

Dada a complexidade e riqueza da linguagem humana, é lógico pensar que os relacionamentos entre as palavras requerem mais do que um espaço bidimensional (como na imagem acima) para serem adequadamente representados. No caso do ChatGPT, ele representa palavras não em um espaço 2D, 3D ou mesmo 4D, mas sim em um espaço de 12.288 dimensões.

Em termos matemáticos, todo esse processo se resume a uma grande matriz. Atribuindo um número único a cada palavra na língua inglesa, podemos representar essa linguagem (codificá-la) em uma matriz de dimensões aproximadas de '2048 x 50.000'. O número 2048 representa o tamanho máximo de entrada permitido no ChatGPT, e 50.000 é o número de palavras que ele possui em seu vocabulário.

pelo modelo para processar uma saída adequada.

3.2.2 Atenção no Chatbot

Outro elemento no funcionamento de Modelos de Linguagem Grandes (LLMs) é a “Atenção”, que se refere a um mecanismo que ajuda o modelo a focar em diferentes partes da sequência de entrada (input), com base no elemento sendo processado na saída (output). Originalmente introduzido no contexto da ‘tradução automática neural’, esse mecanismo tornou-se crucial em várias tarefas de processamento de linguagem natural, tais como os LLMs (MCGUIRE, 2023).

O mecanismo de “atenção” funciona atribuindo diferentes ‘pesos’ a diferentes elementos da sequência de entrada com base na sua relevância para o elemento atual sendo processado na sequência de saída. Isso permite que o modelo foque seletivamente em partes específicas da entrada ao gerar cada parte da saída, possibilitando a captura de dependências e relacionamentos nos dados.

No caso dos LLMs, a atenção é especialmente útil para lidar com o contexto de maneira mais sofisticada. Por exemplo, se perguntarmos a um chatbot quantas espécies de gatos existem, pode ajudá-lo a responder corretamente à pergunta (e de maneira semelhante à humana) se ele der mais importância às palavras ‘gato’ e ‘espécies’ do que às outras palavras.

A “atenção” então ajuda o modelo a atribuir melhores pesos a essas palavras importantes, permitindo que ele leve em consideração o contexto relevante durante as fases de treinamento e geração.

No mecanismo de atenção, cada palavra (ou token) na sequência de entrada está associada a três vetores:

- **Chave (K)**: Uma representação do conteúdo da palavra, usada para comparação com outras palavras.
- **Consulta (C)**: Uma representação da palavra à qual se está prestando atenção, usada para determinar a relevância de outras palavras.
- **Valor (V)**: Uma representação do conteúdo da palavra, usada para contribuir para a saída.

Para calcular o peso da “atenção” para cada palavra na sequência de entrada, uma ‘função de pontuação’ é aplicada aos vetores ‘Consulta’ e ‘Chave’. Essa função de pontuação mede o quão bem esses vetores correspondem para cada palavra na sequência. Funções de pontuação comuns incluem: produto escalar, o produto escalar dimensionado ou uma função aprendida (por exemplo, uma camada de rede neural).

Os resultados da pontuação são normalizados usando uma ‘função softmax’ para obter os pesos da atenção”. Esses pesos indicam quanto foco o modelo deve dar a cada palavra

na sequência de entrada ao gerar a próxima palavra na saída. Os pesos da atenção” são então usados para calcular a soma ponderada dos vetores ‘Valor’ para todas as palavras na sequência de entrada. Essa soma ponderada representa o contexto ou a informação na qual o modelo deve se concentrar ao gerar a próxima palavra.

Essa soma ponderada (informação contextual) é então combinada com os embeddings das palavras de saída previamente geradas (se houver) para formar o contexto para gerar a palavra de saída atual.

O modelo então gera a próxima palavra na sequência de saída com base no contexto e em seu estado interno. Esse processo é repetido iterativamente para gerar uma sequência de palavras.

Em resumo, o “mecanismo de atenção” permite que o LLM aloque dinamicamente seu foco em diferentes partes na sequência de entrada, dando maior ‘peso’ às palavras mais relevantes para gerar a palavra de saída atual. Isso permite que o modelo gere respostas coerentes e contextualmente relevantes, considerando efetivamente as informações de entrada.

3.3 FUNCIONAMENTO DE UM CHATBOT

Agora podemos descrever completamente o funcionamento de um chatbot em um nível tecnológico fundamental. Primeiramente, fornecemos ao chatbot um ‘prompt’. Em seguida, ele transforma a última palavra desse ‘prompt’ em números, codifica-a e depois multiplica esse número pelos pesos que ele aprendeu durante seu treinamento, produzindo assim um “embedding”.

Isso nos fornece uma matriz de 12.000 dimensões. Em seguida, percorremos essa matriz e todos os números associados a ela usando os ‘transformadores de atenção’ do mecanismo de atenção. Isso é feito de forma que algumas palavras no ‘prompt’ (incluindo a última palavra) recebam maior atenção em comparação com as outras palavras no restante do ‘prompt’. Isso é crucial para gerar a saída desejada.

Em seguida, ‘normalizamos’ a matriz resultante para torná-la mais semelhante à matriz original (em termos de estrutura), e, depois disso, a alimentamos para a próxima camada de “transformadores de atenção” e repetimos o processo. No caso do ChatGPT, ele possui um total de 96 camadas para realizar esse procedimento.

Após essas atividades, basicamente realizamos o inverso: pegamos a matriz grande que passou pelas camadas de transformadores de atenção” e revertermos o embedding”. Dessa forma, usamos novamente essas 12.000 dimensões para transformar as palavras em números, e no final, obtemos uma única palavra como resultado.

Pode-se imaginar que, se precisarmos realizar esse processo para cada palavra do ‘prompt’, envolvendo 175 bilhões de operações separadas, isso pode demandar um período considerável de tempo para que um “Modelo de Linguagem Grande” possa fornecer uma

resposta. Portanto, é por isso que há um atraso entre fazer uma pergunta no ChatGPT e receber sua resposta.

É importante lembrar que, mesmo com todo o seu impressionante desempenho, em nenhum momento um Chatbot de IA” verdadeiramente ‘compreende’ a natureza das perguntas feitas a eles. Por exemplo, quando perguntamos a um chatbot qual é o sexto elemento da tabela periódica, a Rede Neural” não está realmente pensando na tabela periódica (ela nem sabe o que são átomos ou elementos químicos). Em vez disso, ela determina, com base na distribuição estatística das palavras em seu vasto conjunto de dados de treinamento, qual palavra é mais provável de seguir a sequência. Então, para a sequência de palavras que forma a pergunta "Qual é o sexto elemento na tabela periódica?", a rede neural do Chatbot indica que a próxima palavra razoável é ‘carbono’. Esta resposta foi gerada a partir da análise dos dados de treinamento com base em tudo o que já foi escrito online.

4 IA GENERATIVA E O DESENVOLVIMENTO DE JOGOS

A "Inteligência Artificial Generativa" tem sido um tema popular em tempos recentes. O uso desse novo conjunto de ferramentas de IA, capazes de reproduzir elementos de texto, voz ou gráficos a partir de simples instruções, abriu as portas para uma nova variedade de aplicações empolgantes (BANH; STROBEL, 2023).

Por exemplo, a partir de um simples 'prompt', tornou-se possível escrever partituras com facilidade e até mesmo compor as letras que as acompanham, criar códigos para gráficos vetoriais e animações, e até mesmo desenvolver trabalhos acadêmicos completos.

No entanto, em meio a essa euforia inicial, compreender o verdadeiro estado atual da "IA Generativa" e seu potencial impacto na indústria de videogames pode ser uma tarefa desafiadora.

Será que a "IA Generativa" realmente poderá abrir as portas para uma nova era de ferramentas que tornarão o desenvolvimento de jogos mais fácil, acessível e rápido? Poderá o uso dessas ferramentas permitir que jogos de grande porte sejam criados agora em questão de semanas ou até mesmo dias, em vez de anos?

As respostas para essas perguntas, e a realidade que as envolve, são consideravelmente complexas. Portanto, é fundamental que nos aprofundemos nas características dessas tecnologias, em como elas funcionam, nas questões que as cercam e na sua adequação para a produção, levando em consideração o contexto real do desenvolvimento de jogos.

4.1 O QUE É IA GENERATIVA

Vamos começar pelo básico. Quando mencionamos "IA Generativa", o que isso quer dizer? Podemos considerar a "IA Generativa" como um termo simplificado que se origina do conceito mais técnico de "Aprendizado de Máquina Generativo". Este último, por sua vez, abrange qualquer sistema de IA, geralmente construído com técnicas de "Aprendizado de Máquina" (Machine Learning), cuja função principal é a criação de algum tipo de saída (output) baseado nos seus dados de entrada (input). Por exemplo, o ChatGPT gera texto, o Stable Diffusion e o DALL-E geram imagens, o GitHub Copilot escreve código, o Speechify produz áudio de voz, entre outros (GRIMMELMANN; LEE; COOPER, 2023).

Em cada uma dessas instâncias, esses sistemas recebem uma entrada e, a partir dessa entrada, geram uma saída projetada para um propósito específico. Embora normalmente associemos sistemas de IA à resolução de problemas, essas tecnologias raramente se limitam a gerar respostas; elas frequentemente criam saídas que podem ser usadas de maneiras práticas e criativas pelos seres humanos. Portanto, o termo "IA Generativa" é apropriado, uma vez que esses sistemas produzem saídas de interesse para os usuários.

O termo "IA Generativa" ganhou destaque graças ao trabalho pioneiro de Ian Goodfellow e seus colaboradores em 2014, quando eles introduziram o conceito de "Redes Generativas Adversariais" (GANs) que revolucionou a geração de imagens. As GANs caminham em conjunto com uma outra técnica de "Aprendizado de Máquina" conhecida como "Codificadores Automáticos Variacionais" (VAEs), que têm se destacado na geração de imagens e no processamento de linguagem natural (GOODFELLOW et al., 2014).

Essas duas metodologias, GANs e VAEs, desempenham um papel central no que hoje chamamos de "Inteligência Artificial Generativa". Embora seus métodos de treinamento subjacentes sejam diferentes, elas compartilham a mesma missão: compreender as propriedades e características essenciais dos dados que elas analisam e, em seguida, armazená-los em um formato compacto, denominado "espaço latente" (CHEIGH, 2023).

Quando um sistema consegue processar adequadamente esse "espaço latente", ele adquire a capacidade de decodificá-lo de maneiras novas e fascinantes, preservando as propriedades identificadas nele. Isso, por exemplo, permite a criação de imagens que se assemelham muito às imagens usadas durante o treinamento. Essa habilidade tem sido aplicada na geração de níveis de jogos, recursos artísticos e outros elementos essenciais para jogos e outras aplicações criativas.

Embora esses avanços sejam impressionantes, é importante notar que nenhum desses conceitos é realmente novo. Os recentes avanços nessa área se baseiam em décadas de pesquisa existente em "Aprendizado de Máquina" e criatividade computacional.

A "IA Generativa" é, essencialmente, um termo mais simplificado, que engloba o conceito de "geração de conteúdo procedural", ou mais especificamente, a "geração de conteúdo procedural usando aprendizado de máquina" (PCG-ML). Além disso, vale ressaltar que sua aplicação não se limita apenas aos jogos.

O conceito de sistemas que utilizam inteligência artificial e "Aprendizado de Máquina" para gerar saídas com propósitos criativos é uma ideia amplamente estabelecida, abrangendo desde sistemas criativos computacionais como o Angelina (o sistema de IA que projetou seus próprios jogos), até a geração de níveis em jogos como no caso de "Super Marios Bros" (Nintendo), e até mesmo a criação de armas no jogo "Shooter Galactic Arms Race" (Bithell Games, Lancarr, 2015) (MARTIN, 2013), (GREEN; KHALIFA, 2018), (HASTINGS; GUHA; STANLEY, 2009).

Embora o termo "PCG-ML" (Procedural Content Generation via Machine Learning) tenha sido adotado por alguns círculos de pesquisa científica há alguns anos atrás, o público em geral tende a preferir o termo mais simples "IA Generativa".

A principal razão pela qual a "IA Generativa" tem evoluído tão rapidamente é devido aos recentes avanços em "Aprendizado Profundo" (Deep Learning), que demonstraram melhorias significativas em relação às suas iterações anteriores. A performance de modelos como GPT-4, por exemplo, é notavelmente superior quando comparada com sua própria capacidade um ano atrás. Ao mesmo tempo, o desempenho de modelos como o DALL-E e

o Stable Diffusion também é muito impressionante, evidenciando o vasto potencial dessa tecnologia (EAPEN et al., 2023).

Graças aos avanços nas infraestruturas de "computação em nuvem", agora é muito mais acessível para o público em geral utilizar esses modelos por meio de interfaces baseadas em texto em páginas web e até mesmo em aplicativos móveis. Qualquer pessoa pode facilmente inserir perguntas para o ChatGPT e receber uma resposta, ou dizer ao Midjourney que tipo de imagens deseja que ele crie.

A criação de interfaces simplificadas para esses modelos não apenas os torna mais atraentes para o usuário comum, mas também mantém o mistério em torno de seu funcionamento, fazendo com que eles pareçam consideravelmente mais inteligentes do que realmente são.

4.1.1 Avanços na IA Generativa e o Mercado

Uma grande parte do mistério que envolve o funcionamento da "IA Generativa" é devido à rápida evolução que seu desempenho sofreu em um período relativamente curto. As principais áreas de aplicação de "IA Generativa", incluindo a geração de texto, a conversão de fala em texto, a transformação de texto em fala e a conversão de texto em imagem, todas as quais estavam em desenvolvimento por décadas, experimentaram avanços notáveis recentemente devido a três fatores principais: (LIN et al., 2023), (THOMAS, 2023), (GARG, 2019)

1. Avanços no campo do "Aprendizado Profundo", principalmente no desenvolvimento e treinamento de infraestruturas de "Redes Neurais" em grande escala.
2. Acesso a conjuntos de dados em larga escala, que possibilitaram uma compreensão mais profunda dos elementos que se deseja reproduzir e a capacidade de gerar uma ampla variedade de respostas alinhadas com essas expectativas.
3. O poder de processamento disponível na computação em nuvem. Isso nos permitiu treinar sistemas em grande escala na nuvem e, posteriormente, implantá-los de maneira prontamente acessível em desktops, laptops, smartphones e em outros dispositivos com conexão à internet.

Outro fator por trás dos recentes avanços no desenvolvimento dessa tecnologia é o investimento de recursos financeiros. Nos últimos 10 anos, tem ocorrido uma mudança significativa na forma como a pesquisa e o desenvolvimento de IA estão sendo financiados, com as corporações assumindo a maior parte da responsabilidade devido às enormes quantidades de dinheiro investidas por elas. (CLARK; PERRAULT, 2022)

Isso culminou em uma corrida competitiva, onde empresas de todos os tamanhos, desde as gigantes corporações até as startups mais modestas, aderiram com o objetivo de

desenvolver sistemas baseados nessas tecnologias. Investidores direcionaram bilhões para esse setor, com uma estimativa de cerca de 70 bilhões de dólares investidos por capitalistas de risco em 2021. Embora tenha ocorrido quedas nesses investimentos, chegando a 46 bilhões em 2022, é importante notar que, considerando outras áreas em declínio, tais como Web3, NFTs e VR, é provável que o investimento nessa área continuará.

Esse cenário também levou à criação de inúmeras novas plataformas que não tiveram sucessos, retratações eventuais e até mesmo retrabalhos, criando assim uma mistura de entusiasmo, altas expectativas e, compreensivelmente, alguma apatia e frustração em relação ao estado atual da “IA Generativa”.

Movidos por incentivo financeiro, seja para garantir a liderança no mercado ou simplesmente para atrair investidores, temos visto lançamentos frequentes de novas ferramentas de IA, sistemas, recursos, demonstrações tecnológicas e muito mais, tanto nas redes sociais quanto por meio de comunicados de imprensa. Somente o tempo revelará quais dessas empresas e seus produtos permanecerão sólidos no mercado e quais acabarão em fracasso.

4.1.2 IA Generativa em Jogos

Agora que vimos os conceitos básicos de “IA Generativa”, surge a questão do seu valor no contexto dos jogos. Os sistemas de “IA Generativa” oferecem um potencial extraordinário para impulsionar o desenvolvimento de jogos. Eles podem ser empregados para uma variedade de finalidades, tais como a criação de texturas e sprites para os elementos do jogo, a geração de animações para personagens, a escrita de descrições de itens, o desenvolvimento de missões, a criação de enredos para jogos de RPGs e até mesmo a elaboração de conversas em tempo real com NPCs que sejam relevantes no jogo e capazes de reagir às ações do jogador. (GWERTZMAN; SOSLOW, 2022)

Todas essas possibilidades são alcançáveis para os desenvolvedores utilizando como base as ferramentas disponíveis atualmente. Desenvolvedores independentes, por exemplo, podem utilizar o “Midjourney” para iniciar seu processo de concepção, auxiliando na criação de artes conceituais para seus projetos. Ao mesmo tempo, programadores podem se beneficiar de ferramentas de programação generativa, como o “GitHub Copilot,” para facilitar a escrita de códigos para seus projetos. Nos encontramos então em uma nova era em que a “IA Generativa” tem o poder de transformar fundamentalmente a maneira como jogos são desenvolvidos.

Muitas corporações têm se interessado na ideia de incorporar a “IA Generativa” em seus processos de produção. Esse entusiasmo tem despertado o interesse tanto de desenvolvedores quanto de jogadores.

No entanto, apesar do desejo das empresas de tomarem liderança nesse novo mercado, o caminho para o sucesso nesse área não é uma linha reta. Desenvolver essas ferramentas de forma segura para uso, as adaptando às necessidades dos desenvolvedores de modo que não resultem em constrangimento público ou até mesmo em problemas legais é, por si

só, um desafio significativo. Embora possamos encontrar exemplos nas redes sociais de pessoas provando que é possível usar essas ferramentas para criar jogos, a transição de fazer com que um conceito interessante funcione em um vídeo para um produto comercial completo é um caminho longo, legalmente incerto e repleto de elementos desconhecidos.

Muitas empresas estão trabalhando em resolver esses desafios. Empresas focadas na criação de jogos como a Latitude, os responsáveis pelo “AI Dungeon”, em colaboração com outras empresas como a Hidden Door, estão desenvolvendo modelos de linguagem dedicados à geração de histórias. (LAI, 2023)

A Ubisoft está explorando a aplicação de geração de texto para roteiros, enquanto a Activision Blizzard dispõe de seu próprio conjunto de ferramentas voltadas para a geração de arte (IVAN, 2023), (WRIGHT, 2023) . O estado da arte da “IA Generativa” aplicada a jogos provavelmente passará por uma transformação significativa em 2023 e nos anos subsequentes.

4.2 DESAFIOS NO CAMPO DE IA GENERATIVA

Vejamos agora alguns dos principais desafios no campo da “IA Generativa”. Em primeiro lugar, é importante enfrentar a realidade: embora tenha-se testemunhado melhorias notáveis nos sistemas de “IA Generativa”, muitas vezes essas tecnologias não são tão poderosas ou eficazes quanto elas são anunciadas. Embora haja grandes expectativas de que a “IA Generativa” irá revolucionar a indústria de jogos, ainda estamos a meses (se não anos), de ver muitas dessas tecnologias atingirem um nível de confiabilidade, sustentabilidade, praticidade e perfeição que correspondam às demonstrações tecnológicas exibidas. (WINTERMEYER, 2023)

É importante notar que essa perspectiva é influenciada pelo contexto recente da empolgação em torno de Web3 e NFTs, que também prometeram transformar a indústria de diversas formas. No entanto, ao contrário dos NFTs, a IA tem, de fato, o potencial de alterar significativamente o desenvolvimento de jogos. No entanto, essas tecnologias e as ferramentas e integrações correspondentes estão evoluindo a diferentes velocidades ao redor do mundo, à medida que várias empresas enfrentam o desafio de tornar essa tecnologia mais acessível, sendo que algumas delas começaram com vantagens nesse campo.

No entanto, isso é apenas o começo; há um longo caminho a ser percorrido desde uma demonstração tecnológica interessante até o estabelecimento de uma cadeia de ferramentas pronta para uso por desenvolvedores. É igualmente crucial submeter essas novas técnicas a uma rigorosa avaliação quanto ao seu realismo. Certamente, uma ferramenta pode gerar diálogos envolventes, texturas visualmente atraentes e até mesmo NPCs ou níveis de jogo. Mas a questão crucial é: ela consegue produzir resultados de alta qualidade de forma consistente e confiável 100% do tempo? Se persistirem problemas e riscos que demandem intervenção humana, a eficácia dessas ferramentas é limitada, especialmente

se forem promovidas como soluções autônomas, sem a necessidade de envolvimento do usuário.

Além disso, é importante destacar que muitas ferramentas de “IA Generativa” devem ser capazes de abordar desafios que os desenvolvedores de jogos enfrentam na prática e atender às operações dos estúdios de jogos. Os jogos, desde projetos independentes pequenos até títulos grandes e famosos, são empreendimentos complexos envolvendo várias pessoas que desempenham diversas funções nesse projeto.

Portanto, o papel da “IA Generativa” deve ser o de resolver eficientemente os problemas enfrentados pelos desenvolvedores de jogos e aumentar a sua produtividade, em vez de buscar substituí-los. Uma das principais razões pelas quais a adoção do “Aprendizado de Máquina” no desenvolvimento de videogames demorou tanto foi a incompatibilidade das capacidades iniciais do mesmo de lidar com os desafios reais enfrentados pelos desenvolvedores de jogos. Somente agora, com técnicas de “Aprendizado de Máquina” capazes de solucionar eficazmente os desafios de produção enfrentados pelos desenvolvedores, essa tecnologia obteve uma relevância totalmente nova. (DEVELOPERS, 2023)

4.2.1 Problemas Legais no Uso de IA Generativa

Focando nos aspectos práticos, um dos maiores desafios que afetam as ferramentas de “IA Generativa” (não apenas na indústria de jogos, mas em todo o campo) é em relação à base de dados usada para treiná-las. Geradores de texto exigem uma quantidade substancial de dados de texto para aprender como interpretá-los, desenvolver suas habilidades de generalização e, eventualmente, serem capazes de produzir seus próprios textos. De maneira análoga, geradores de imagens necessitam de amplo acesso a conjuntos extensos de dados de imagens para aprender e, subsequentemente, gerar imagens. No entanto, o problema crucial se resume a de onde provêm esses dados e, igualmente importante, as pessoas que contribuíram com esses dados deram permissão para sua utilização? (MORARTY, 2023)

Uma boa parte das ferramentas de “IA Generativa” atuais utiliza dados de treinamento que não são necessariamente de domínio público e, em alguns casos, a composição exata desses conjuntos de dados não é divulgada de maneira transparente. Isso gera preocupações significativas para os usuários desses sistemas.

Primeiramente, a falta de transparência em relação ao conteúdo desse conjunto de dados de treinamento pode resultar em resultados desconhecidos e, conseqüentemente, uma falta de confiança nos sistemas que as utilizam. Afinal, as ferramentas de “IA Generativa” são essencialmente imitadores altamente complexos. Um gerador de texto imita o texto que foi usado para treiná-lo, e um gerador de imagens imita a arte presente na sua base de treinamento. É compreensível que um usuário deseje ter a certeza de que a IA que ele está empregando foi treinada com dados relevantes para garantir resultados práticos e evitar saídas indesejadas. Por exemplo, a aparição de um NPC que proclama discursos

racistas, sexistas, homofóbicos ou transfóbicos devido ao seu aprendizado em um conjunto de treinamento problemático não seria aceitável no desenvolvimento de um jogo famoso.

Segundo, agravando os problemas relacionados aos dados e ao acesso, existem também implicações legais significativas relacionadas a essas ferramentas de “IA Generativa”, sendo a origem dos dados utilizados uma questão crucial. Foi descoberto que ferramentas de geração de imagens incluem em seus conjuntos de dados obras de artistas que não autorizaram sua utilização. Pesquisas demonstraram maneiras eficazes de extrair imagens quase idênticas às do conjunto de treinamento, acentuando ainda mais a questão da responsabilidade desses sistemas. Isso gerou uma reação significativa nas comunidades de arte online, com preocupações sobre a extração não autorizada de arte por empresas, bem como um movimento de boicote à arte gerada por IA. Essas preocupações não se limitam ao âmbito das artes; há também um caso legal em andamento envolvendo a Microsoft, GitHub e OpenAI, que suscita questões sobre como as ferramentas de geração de código aprenderam a programar, possivelmente violando licenças de código-fonte protegidas por direitos autorais. Isso tem implicações significativas para o futuro da aquisição de dados para sistemas de “IA Generativa”.

Isso leva a questões complexas relacionadas a direitos autorais e uso justo. Se a arte gerada por IA for incorporada em partes de um jogo, tais como arte conceitual, designs de personagens, texturas e outros elementos, e for descoberto que as mesmas se originaram de um artista que não deu seu consentimento, isso pode acarretar implicações significativas, não apenas para o criador do gerador, mas também para qualquer arte derivada dele. Estúdios grandes podem hesitar em querer enfrentar as possíveis consequências legais, a menos que possam garantir que estão utilizando apenas dados de treinamento internos, uma abordagem que muitos estúdios estão atualmente explorando.

No entanto, essa abordagem não resolve completamente o problema da propriedade dos produtos gerados por esses sistemas. Atualmente, não existe base legal para conferir às saídas produzidas por uma “IA Generativa” o status de direitos autorais. O fato de alguém instruir um gerador como ChatGPT a escrever algo não significa que essa pessoa detenha os direitos autorais sobre a resposta gerada, e o mesmo princípio se aplica a outras formas de conteúdo criado por “IA Generativa”. Em vez disso, os elementos gerados estão disponíveis para uso por qualquer pessoa, o que pode limitar sua utilidade em ambientes de produção onde a proteção de direitos autorais é importante.

4.2.2 Como IA Generativa Pode Afetar os Desenvolvedores

Outro grande problema no campo da “IA Generativa” que é frequentemente negligenciado é a importância do elemento humano. As pessoas que criam jogos são a fonte da criatividade desses jogos, e a “IA Generativa” precisa ser construída de modo a apoiar essa criatividade. Grande parte da discussão em torno da “IA Generativa” tem se concentrado na ideia de usar a IA para substituir o trabalho humano, alimentando a narrativa

de que a IA irá acabar efetivamente com vários papéis das pessoas no processo criativo, desde artistas de textura até artistas conceituais, modeladores 3D, animadores, designers de som, dubladores e até programadores. Isso é um exemplo clássico de capitalismo conservador que promove a automação de empregos onde as pessoas não apenas dependem para seu sustento, mas também têm um forte apego pessoal ao seu trabalho. No entanto, essa narrativa ignora o fato de que a “IA Generativa” funciona melhor quando ela está apoiando a criatividade humana, não quando a substitui. (WU et al., 2021)

Embora seja verdade que a “IA Generativa” mudará a forma como muitos empregos e funções são realizados em diversos setores, ela deverá fazer isso aprimorando as capacidades humanas em vez de substituí-las completamente. Alguns empregos podem ser afetados, e algumas equipes podem ser reduzidas, mas advogar pela substituição total dos desenvolvedores por IA não é uma boa ideia. Os jogos são uma forma de expressão criativa, e os desenvolvedores precisam estar em seu auge criativo para produzir os melhores jogos possíveis. A IA pode certamente apoiar e facilitar esse processo, mas seu papel principal deve ser o de ajudar os criadores a atingir seus objetivos criativos, não substituí-los.

Existem várias aplicações de “Aprendizado de Máquina” que se concentram em ajudar os desenvolvedores de jogos em seus processos, ao invés de tentar substituí-los, tais como a melhoria de texturas, a correspondência de movimento e a detecção de trapaças. Essas aplicações abordam desafios de produção de maneira mais rápida e eficaz do que aumentar a força de trabalho.

A rejeição à “IA Generativa” muitas vezes decorre da percepção de que ela irá substituir indivíduos nos campos criativos em vez de apoiá-los em seus trabalhos. Mesmo para empresas que buscam automação completa, a presença de seres humanos no processo criativo permanece essencial para que o mesmo funcione eficientemente. A “IA Generativa” deve ser vista como uma ferramenta que permite aos seres humanos desempenhar seus trabalhos com mais eficiência, em vez de ser uma forma de substituí-los.

5 USO DE LLMS E IA GENERATIVA EM JOGOS

Vimos que a "IA Generativa" é uma tecnologia poderosa com uma ampla variedade de aplicações, estando na vanguarda da inovação em diversas indústrias e possuindo o potencial de revolucionar a maneira como criamos, interagimos e experimentamos conteúdos em jogos.

A "IA Generativa" em jogos não é apenas uma ferramenta para a produção de conteúdo; ela também pode ser uma parte integral da experiência e jogabilidade. Vimos que, tradicionalmente, a IA em jogos tem se limitado ao controle de "Personagens Não-Jogáveis" (NPCs) ou à geração de eventos roteirizados. No entanto, a "IA Generativa" tem o potencial de permitir uma experiência mais dinâmica, ao possibilitar que os jogadores se envolvam com os mundos fictícios dos jogos e moldem ativamente as narrativas dos mesmos.

Isso pode ser feito usando a "IA Generativa" para fornecer uma estrutura que capacite os jogadores a co-criarem as histórias com as quais estão interagindo. Ao contrário dos jogos convencionais, que possuem narrativas predefinidas, a "IA Generativa" é capaz de se adaptar às escolhas e ações dos jogadores, oferecendo um nível de flexibilidade e interatividade que anteriormente não era possível. Ela pode então servir como um parceiro colaborativo, estabelecendo limites, apresentando desafios e respondendo às decisões dos jogadores em tempo real.

Uma distinção importante a ser feita sobre a "IA Generativa" nos jogos é o seu papel em estimular a criatividade. Enquanto sistemas de IA excelam em prever resultados e apresentar um espectro variado de possibilidades, eles não possuem criatividade em si. Em vez disso, eles facilitam e aprimoram os processos criativos dos jogadores humanos. Logo, teríamos um desempenho melhor nos concentrando na energia espontânea e improvisada que surge da interação entre os jogadores e a IA.

A tecnologia de "IA Generativa" pode servir como um catalisador para a narrativa colaborativa, oferecendo uma experiência rica e imersiva e permitindo que os jogadores explorem e interajam com mundos fictícios, proporcionando uma nova mistura única de fanfiction e elementos de interpretação de personagens. Um exemplo de onde ela pode ser usada é em RPGs de mesa (TTRPGs), onde, ao empregar um sistema de "IA Generativa" no plano de fundo para capacitar as habilidades narrativas dos jogadores, pode-se eliminar muitas das barreiras logísticas associadas a esse gênero de jogos e torná-lo mais acessível a um público mais amplo. (SANTIAGO et al., 2023)

O principal propósito da "IA Generativa" em jogos é o seu papel como um facilitador e ferramenta, em vez de ser a força criativa primária do jogo. Ele serve como um meio para permitir uma narrativa colaborativa, onde os jogadores podem interagir entre si e com o próprio sistema, gerando momentos e experiências mais memoráveis.

Os jogadores seriam então incentivados a trocar ideias entre si e responder às sugestões do sistema, criando uma narrativa dinâmica e envolvente. Por exemplo, ao enfrentar um adversário difícil, um jogador pode escolher receber ajuda no combate, enquanto outro pode optar por distrair o inimigo. O sistema poderia então levar em consideração essas ações e intenções, incorporando-as na narrativa, resultando em desdobramentos surpreendentes e interessantes.

Temos então que a "IA Generativa" em jogos pode representar uma mudança significativa na área de jogos, indo de um consumo passivo de conteúdo para uma participação ativa em sua criação. Ela pode oferecer um vislumbre de um futuro onde a tecnologia aumenta a criatividade humana, tornando a narrativa mais envolvente e personalizada, ao mesmo tempo em que democratiza o prazer dessas experiências. Conforme essa tecnologia continua a evoluir, prevê-se que seu impacto na indústria de jogos será imenso.

5.1 DESAFIOS DE DESIGN

Um aspecto essencial no design de jogos é a incorporação de restrições. Jogos se baseiam em um conjunto de regras e limitações que moldam a experiência do jogador, garantindo que o jogo permaneça envolvente e agradável. Temos então que restrições são necessárias para manter parte do fator "diversão".

Jogos são, essencialmente, sistemas intrincados de regras e restrições que fornecem estrutura e orientação aos seus jogadores. Embora o uso de "IA Generativa" possa levar alguém a imaginar uma liberdade sem limites, é crucial estabelecer um equilíbrio entre a agência do jogador e a manutenção de uma experiência de jogo coerente e agradável. (BEDINGFIELD, 2023)

O desafio então é introduzir estrutura e regras sem sufocar a criatividade e a agência. O objetivo é empregar "IA Generativa" para criar um jogo genuíno, em vez de tê-la como uma mera ferramenta de escrita ou improvisação. Ao contrário de ferramentas de escrita ou plataformas de improvisação, os jogos envolvem um elemento de risco e a possibilidade de falha. Os jogadores devem tomar ações, fazer escolhas e enfrentar consequências, incluindo a possibilidade de perder.

Portanto, temos que o desafio de design é definir os limites e a estrutura do jogo. Mas como um sistema que permite aos jogadores tomarem qualquer ação desejada dentro de um mundo responsivo pode também fornecer estrutura e manter a integridade da narrativa? Uma possível resposta pode estar em se inspirar nas dinâmicas entre GMs (Game Masters) em TTRPGs, que estabelecem convenções de narrativa social para direcionar a história e acomodar as preferências dos jogadores. (SANTIAGO et al., 2023)

Por exemplo, o sistema pode incorporar convenções do tipo "não separar o grupo" ou proibir comportamentos disruptivos dentro da narrativa. Assim, enquanto ele permite uma ampla gama de escolhas para os jogadores, ele também respeita certas restrições

que ajudam a preservar a experiência geral. Levar em consideração os elementos sociais dentro dos jogos adiciona outra camada de complexidade. A possibilidade de os jogadores serem amigos e interagirem fora do sistema é outra variável a levar em conta. Ainda assim, é essencial projetar o sistema para se adaptar a vários comportamentos dos jogadores, incluindo os disruptivos ou não cooperativos, enquanto o mesmo procura manter a integridade e a diversão do jogo.

Incorporar "IA Generativa" no design de um jogo envolve um equilíbrio delicado entre estabelecer restrições e permitir a agência do jogador. Esse é um problema de design complexo que requer uma consideração cuidadosa de como orientar as ações do jogador, ao mesmo tempo em que se procura preservar a sua liberdade de explorar e moldar a narrativa dentro dos limites de uma experiência de jogo envolvente.

5.1.1 Limites Tecnológicos

Na área de jogos impulsionados por "IA Generativa", os obstáculos tecnológicos frequentemente se concentram em manter o controle sobre a narrativa, garantir a segurança do jogador e manter uma experiência de jogo estruturada. Esses desafios são intrínsecos ao utilizarmos "Modelos de Linguagem Grande" (LLM, como o ChatGPT) para contar histórias interativas.

Um desafio significativo reside na controlabilidade. Ao empregarmos "Modelos de Linguagem Grande" para contar histórias dinâmicas e interativas, esses modelos podem se comportar de maneira imprevisível. Eles podem ter dificuldade em manter a consistência, esquecer o contexto, introduzir inconsistências ou desviar a narrativa de seu caminho pretendido. Essa falta de controle pode resultar em uma experiência de jogo menos satisfatória. (RABBIT, 2023)

Possíveis soluções para esse problema envolvem treinar o "LLM" em um conjunto de dados de texto e código cuidadosamente selecionados e relevantes para o mundo do jogo e a história em questão; isso ajudará o "LLM" a gerar conteúdo consistente com o ambiente e as regras do jogo. Também podemos empregar técnicas como "engenharia de prompts" e "aprendizado por reforço" para orientar a saída do LLM; isso pode ajudar o LLM a gerar conteúdo mais alinhado com as intenções do desenvolvedor.

Outra restrição a ser abordada é a questão de segurança e o controle de conteúdo. Os "Modelos de Linguagem Grande" às vezes podem gerar conteúdo errático ou problemático, o que pode ser prejudicial ou ofensivo para os jogadores. Garantir que o jogo permaneça seguro e respeitoso é um elemento importante na incorporação desses sistemas. (FER-RARA, 2023)

Para combater esse problema, podemos usar técnicas como filtragem e moderação para remover conteúdo problemático; isso pode envolver o uso de moderadores humanos ou treinar um "LLM" separadamente para identificar e sinalizar conteúdo prejudicial. Além disso, o sistema pode ser treinado para interpretar entradas de texto e garantir que

certos termos ou palavras sensíveis não sejam permitidos, impedindo que os jogadores introduzam elementos inapropriados nele.

Criar conteúdo dinâmico que responda às ações do jogador ao mesmo tempo em que mantém uma experiência de jogo estruturada e coerente é o principal desafio. É necessário garantir que o mundo do jogo evolua de forma lógica e siga uma estrutura bem definida.

Possíveis soluções para esse desafio envolvem o desenvolvimento de algoritmos e técnicas para gerar conteúdo estruturado, tais como usar o "aprendizado por reforço" para treinar o LLM a gerar conteúdo consistente com a estrutura e as regras do jogo.

5.2 UTILIZANDO LLMS EM JOGOS

Um aspecto intrigante ao utilizar "Modelos de Linguagem Grande" (LLMs) no contexto da "IA Generativa" em jogos é a capacidade de misturar estruturas de dados simples com linguagem, proporcionando assim uma forma de operar em ambos. Fornecer essa tecnologia a escritores e desenvolvedores pode permitir que eles criem seus próprios mundos de forma interativa e imersiva. No entanto, surge um desafio significativo: como garantir que os LLMs capturem efetivamente os temas principais e nuances desses diversos universos fictícios.

Um dos principais desafios é a dificuldade de fazer o modelo de linguagem destacar os temas e características principais de um mundo específico, como o tom narrativo. Por exemplo, "Game of Thrones" é marcado por violência e intriga, enquanto "Star Trek" gira em torno de otimismo, exploração e resolução de problemas. Adaptar um "LLM" para representar fielmente esses diversos elementos temáticos pode ser uma tarefa complexa. (PORTER, 2017)

Uma possível solução envolve fornecer um meio estruturado para definir a natureza de um mundo, como usar subgêneros para categorizar e especificar a essência de um mundo. Por exemplo, um mundo pode ser definido como uma mistura de comédia, fantasia, drama ou romance. As possíveis porcentagens de subgênero que o trabalho em questão possui podem então servir como base para estabelecer as regras de como o sistema deve se comportar dentro da narrativa. Elas determinam a frequência e o estilo de vários elementos na história, como assassinato, ação, dramas baseados em relacionamentos e narrativas dramáticas.

Outro problema é como o modelo de linguagem pode expressar eventos e ações, como um personagem perdendo HP após ser atacado. Isso deve ser representado como um confronto dramático em estilo régio ou como uma luta de boxe esportiva? O desafio está em determinar os "pesos" apropriados para a expressão linguística e visual, de modo que eles estejam alinhados com os temas definidos no trabalho fictício em questão.

Essas soluções envolvem reduzir as possibilidades infinitas inicialmente oferecidas pelo modelo de linguagem para um conjunto gerenciável de alterações e parâmetros do me-

canismo do jogo. Esses são então expandidos novamente em possibilidades de expressão infinitas, criando uma ilusão de um espaço narrativo vasto e imersivo. Essa abordagem permite que o modelo de linguagem se adapte e responda dinamicamente às demandas temáticas e tonais dos mundos fictícios que ele dá vida.

5.2.1 Treinamento de LLMs

Treinar um modelo de linguagem para capturar e usar palavras (ou um vocabulário específico) a fim de criar um mundo distinto dentro de um jogo também apresenta seus desafios. Esses problemas são abordados combinando as capacidades de modelagem de linguagem com o conhecimento específico do domínio em questão.

O problema descrito acima envolve identificar e usar as palavras e frases que são únicas para um mundo fictício específico. Por exemplo, no universo de "Star Wars", termos como "a Força" são fundamentais para a história e aparecem com frequência. No entanto, além desses termos característicos, o mundo pode abranger várias tramas narrativas, como romance, mistério ou ação. O modelo de linguagem deve então ser treinado não apenas para reconhecer esses termos distintivos, mas também para usá-los apropriadamente de modo a imergir os jogadores na atmosfera desse mundo.

Uma abordagem possível para esse problema é gerenciar o vocabulário e os padrões de linguagem específicos do mundo. Esse processo de gerenciamento envolve analisar os textos existentes sobre o universo em questão e extrair palavras e frases relevantes que contribuam para a singularidade do mundo. No entanto, isso não termina com a extração; um processo de revisão e aprovação manual também é necessário. Os desenvolvedores de jogos devem confirmar que palavras e frases específicas estão sendo usadas apropriadamente no contexto do jogo.

Além de um vocabulário específico, o processo de treinamento também explora o uso de "tropes" (tropes literárias são recursos de narração, ou seja, formas que guiam as histórias) e elementos únicos que definem o mundo. Isso inclui coisas como o tipo de personagens, suas características (humanos, alienígenas, etc.) e a presença de tecnologias ou fenômenos específicos (por exemplo, viagem mais rápida que a luz). Reconhecer e incorporar esses elementos na narrativa do jogo é essencial para manter a autenticidade dentro dele.

Outra dimensão desse desafio é a necessidade de personalizar as saídas do modelo de linguagem com base nas preferências do autor do mundo do jogo. Autores podem ter preferências distintas em mente para seus mundos, e o modelo de linguagem deve ser adaptável para acomodar essas variações. Essa personalização permite que autores definam aspectos únicos, como a forma como os personagens reagem a eventos ou os efeitos específicos de ações dentro de seu mundo.

5.2.2 Modelos Grandes e Pequenos

No campo da "IA Generativa" para jogos, uma interação interessante entre tecnologia e viabilidade econômica está modelando o cenário da narrativa interativa. Vamos então explorar as vantagens e nuances de empregar "LLMs", tanto os modelos menores quanto os modelos maiores, no contexto do desenvolvimento de jogos. (TEAM, 2023a), (TEAM, 2023b)

Os modelos menores são caracterizados por sua natureza focada e especializada. Esses servidores são projetados para lidar com tarefas ou funcionalidades específicas dentro de um sistema de jogo de "IA Generativa". Alguns aspectos-chave desses modelos menores e servidores associados são:

- **Funcionalidade Especializada:** Modelos menores são projetados para realizar funções específicas na experiência de jogo. Por exemplo, eles podem se destacar na geração de pontos para o enredo, interações entre personagens ou elementos narrativos. Essa especialização garante que cada modelo desempenhe sua tarefa designada excepcionalmente bem.
- **Abordagem de Conjunto:** Em um sistema de "IA Generativa" que utiliza modelos menores, pode-se adotar uma abordagem de conjunto. Isso significa que vários modelos menores trabalham em conjunto para contribuir com diversos aspectos do jogo. Cada modelo é responsável por uma função específica, como geração de diálogo ou comportamento de personagens.
- **Integração de Dados Estruturados:** Modelos menores são hábeis em incorporar dados estruturados na narrativa do jogo. Eles podem transformar dados estruturados do jogo em elementos de história significativos, garantindo que o jogo permaneça controlável e esteja alinhado com a funcionalidade geral do mecanismo de jogo.
- **Eficiência de Custo:** Modelos menores são geralmente mais eficientes em termos de recursos computacionais. Eles são usados estrategicamente em conjunto com outras técnicas, como pré-geração e classificação baseada em CPU, para otimizar a utilização de recursos. Essa eficiência de custos é vantajosa tanto para o desenvolvimento quanto para a implantação.

Modelos grandes, no contexto da "IA Generativa", referem-se a modelos de linguagem massivos (como GPT-3, GPT-4 e além). Esses modelos são caracterizados por seu tamanho vasto, que é geralmente medido em termos do seu número de parâmetros. Alguns dos aspectos-chave e vantagens de modelos maiores são:

- **Geração de Linguagem:** Modelos maiores (como o GPT-3) são conhecidos por sua incrível habilidade de gerar texto semelhante ao humano. Eles foram pré-treinados

em grandes quantidades de dados textuais, o que lhes confere um profundo entendimento da linguagem e de contexto. Isso permite que eles produzam narrativas, diálogos e descrições coerentes e contextualmente relevantes.

- **Saídas de Alta Qualidade:** Jogos que utilizam modelos maiores se beneficiam de saídas de texto de alta qualidade. Isso se traduz em narrativas envolventes, interações realistas entre personagens e ambientes de jogo imersivos. Os jogadores podem experimentar narrativas ricas e jogabilidade dinâmica graças à sofisticação desses modelos.
- **Referência e Benchmarking:** Mesmo que um sistema de "IA Generativa" dependa principalmente de modelos menores, a comparação com modelos maiores serve como um valioso ponto de referência. Isso permite que os desenvolvedores meçam o desempenho de seu sistema em termos de qualidade de geração de linguagem em relação a um padrão forte.
- **Intensivo em Recursos:** É importante observar que implantar e manter modelos maiores pode ser intensivo em recursos, tanto em termos de potência computacional quanto de custo. Esses modelos requerem recursos computacionais substanciais para funcionar de maneira eficiente, tornando-os adequados para projetos bem financiados.

Portanto, modelos menores fornecem especialização, flexibilidade, integração de dados estruturados, eficiência de custos e adaptabilidade para a "IA Generativa" em jogos. Por outro lado, modelos maiores oferecem desempenho de ponta e servem como padrões de referência no campo. A escolha entre modelos menores e maiores depende dos requisitos em questão, objetivos e considerações econômicas da experiência de jogo, equilibrando o desempenho com a eficiência de custos.

5.3 INTELIGÊNCIA ARTIFICIAL GERAL (AGI)

O termo AGI, que significa "Artificial General Intelligence" (Inteligência Artificial Geral), refere-se a sistemas autônomos e inteligentes que possuem a capacidade de aprender e realizar uma ampla variedade de tarefas.

O conceito de "AGI" engloba a ideia de criar máquinas com inteligência semelhante à humana, capazes de raciocinar, aprender e resolver problemas em diversos campos diferentes. No entanto, chegar a uma definição precisa de inteligência continua sendo um desafio, sendo este um grande objeto de debate no campo de IA.

O "Teste de Turing", que testa a capacidade de um computador de exibir comportamento inteligente equivalente ao de um ser humano, foi inicialmente visto como o ponto

de referência para alcançar a inteligência artificial. Ele envolve a capacidade de uma máquina de participar de conversas em linguagem natural indistinguíveis de um ser humano. No entanto, esse ponto de referência evoluiu ao longo do tempo, e argumenta-se que ele não abrange todo o espectro da inteligência. (SULLIVAN, 2023)

A discussão sobre "AGI" leva a uma reavaliação do próprio conceito de inteligência. O surgimento de modelos de linguagem altamente avançados, tais como o ChatGPT, levantou questões sobre a natureza da inteligência, da criatividade e dos limites das capacidades da IA.

O desenvolvimento da "AGI" apresenta uma incerteza significativa. Embora os avanços em IA e "Aprendizado de Máquina" tenham levado a realizações impressionantes, "AGI" continua sendo um objetivo complexo e elusivo.

A busca por "AGI" resulta em profundas discussões éticas, filosóficas e sociais. Questões sobre o que constitui uma verdadeira inteligência, as responsabilidades associadas ao desenvolvimento da "AGI" e as potenciais consequências do surgimento da "AGI" tomam o centro desses debates.

Há uma preocupação de que o intenso foco na criação de uma "AGI" e nos riscos relacionados a ela possa estar desviando a atenção para outros aspectos do desenvolvimento de IA, como sistemas de IA e "Modelos de Linguagem Grandes".

Embora "AGI" continue sendo um objetivo a longo prazo para o campo de IA, ainda há uma necessidade de abordar as implicações éticas e práticas das tecnologias de IA em suas formas atuais, incluindo seu potencial para amplificar preconceitos e causar danos. Equilibrar a busca por "AGI" com o desenvolvimento responsável da IA é importante.

5.4 COMPORTAMENTO EMERGENTE

O fenômeno de comportamento emergente reside na interseção entre o design de jogos e a inteligência artificial (IA). Este é um conceito enraizado no mundo dos jogos e está cada vez mais relevante no campo das tecnologias impulsionadas pela IA.

Na área de jogos, o comportamento emergente refere-se a ações e interações imprevisíveis e geradas pelos jogadores, que surgem das interações entre eles e as regras e mecânicas do jogo. Imagine um jogo de "sandbox" onde os jogadores têm liberdade para experimentar e criar experiências únicas. Nesses jogos, os jogadores frequentemente surpreendem os desenvolvedores de jogos ao descobrir estratégias inovadoras ou maneiras criativas de interagir com o ambiente do jogo. Esses comportamentos imprevistos dos jogadores são a marca registrada do comportamento emergente no mundo dos jogos.

Esse conceito de "comportamento emergente" encontra um paralelo em IA, especialmente em "Modelos de Linguagem Grandes" como o GPT-3 e outros. Inicialmente, esses modelos de IA podem parecer simples, fornecendo respostas que se alinham com a entrada que eles recebem. No entanto, à medida que esses modelos aumentam em tamanho

e complexidade, algo extraordinário acontece. Eles começam a exibir comportamentos inesperados. (WEI et al., 2022)

Imagine ter uma conversa com um modelo de linguagem. Inicialmente, ele pode produzir respostas padrões e esperadas. Mas, à medida que a capacidade do modelo aumenta, ele começa a surpreender os usuários e desenvolvedores com sua capacidade de respostas criativas (e às vezes, alucinatórias). Esses comportamentos inesperados não são sempre considerados falhas; em vez disso, eles se tornam características valiosas para determinadas aplicações.

A integração de sistemas de IA em jogos pode adicionar uma nova camada ao conceito de comportamento emergente. Agora, designers de jogos têm a oportunidade de aproveitar a imprevisibilidade e a criatividade da IA para melhorar a jogabilidade. Assim como os jogadores podem descobrir novas dimensões dos jogos por meio de suas interações, os sistemas de IA trazem suas próprias qualidades únicas para a experiência de jogo.

Projetar jogos com o "comportamento emergente" em mente significa abraçar a ideia de que os jogadores se envolverão com o jogo de maneiras difíceis de prever. Trata-se de criar um ambiente dinâmico onde a IA e os jogadores coexistam, gerando experiências inovadoras que borraram a linha entre a jogabilidade scriptada e a inovação impulsionada pelo jogador.

O "aprendizado por reforço", uma técnica fundamental de IA, desempenha um papel crucial na formação do "comportamento emergente". Agentes de IA, assim como jogadores em um jogo, tomam decisões com base nas informações disponíveis para eles e recebem feedback com base nessas decisões. Esse processo de aprendizado e adaptação reflete a maneira como os jogadores navegam pelos desafios de um jogo.

Os designers de jogos são especialistas em trabalhar com sistemas complexos, onde regras aparentemente simples podem levar a comportamentos intrincados e inesperados por parte dos jogadores. Essa perícia nem sempre é reconhecida fora do mundo dos jogos, mas é inestimável ao integrar sistemas de IA em jogos.

O comportamento emergente não se limita apenas aos jogos; ele se estende a muitas aplicações de IA. Assim como os jogadores moldam suas experiências de jogo, modelos de IA têm o potencial de influenciar diversas áreas, desde a geração de conteúdo criativo até a tomada de decisões.

Esta é uma fronteira empolgante, onde modelos de IA se tornam co-criadores dinâmicos, enriquecendo o mundo dos jogos e abrindo novas possibilidades em diversas indústrias. À medida que continuamos a explorar o comportamento emergente, descobrimos o poder da colaboração entre a inventividade humana e a inteligência artificial.

6 SIMULADOR DE TTRPG - PARTE PRÁTICA

O objetivo da parte prática deste projeto é criar uma aplicação web que simule uma experiência de RPGs de Mesa (TTRPG, na sigla em inglês) empregando LLMs e IA Generativa. TTRPGs são conhecidos por suas histórias interativas e narrativas impulsionadas pelos jogadores, e esta aplicação web busca capturar a essência dessa experiência em formato digital. Ao combinar as entradas do usuário com as capacidades de “Modelos de Linguagem Grandes” (LLM), o projeto tem como objetivo fornecer aos jogadores os meios para criar uma narrativa única e em constante evolução com base em suas escolhas.

No contexto dos TTRPGs, nos quais os jogadores geralmente criam colaborativamente uma história enquanto são orientados por um Game Master (Mestre de Jogo), esta aplicação web atua como um Game Master digital, gerando descrições de personagens, detalhes do mundo e eventos de história dinamicamente. O objetivo é oferecer aos jogadores uma experiência contínua e imersiva, permitindo que moldem suas próprias aventuras por meio de entrada criativa e tomada de decisões.

Tecnologicamente, o projeto utiliza o React, uma biblioteca popular de JavaScript para construir interfaces de usuário, a fim de criar uma interface responsiva e interativa. Além disso, a aplicação se integra a modelos de linguagem externos por meio de solicitações de API para gerar os detalhes dos personagens e do mundo, bem como avançar dinamicamente a história com base nas escolhas dos jogadores. Essa combinação de tecnologias modernas de desenvolvimento web e processamento de linguagem natural possibilita uma simulação fluida e interessante de TTRPG.

6.1 VISÃO GERAL DO PROJETO

Esta aplicação web de Simulação de RPG engloba diversas funcionalidades projetadas para proporcionar aos usuários uma experiência abrangente e imersiva. As principais características podem ser delineadas da seguinte forma:

Criação de Personagem: Os usuários iniciam sua jornada fornecendo um nome e uma descrição para o personagem. A aplicação utiliza Modelos de Linguagem externos por meio de solicitações de API para gerar uma imagem correspondente à descrição do personagem. Este passo garante uma criação de um personagem visualmente personalizado e único, de acordo com a entrada do usuário.

Construção de Mundo: Após a criação do personagem, os usuários então descrevem o mundo em que sua história se passa. Semelhante à criação de personagens, a aplicação utiliza Modelos de Linguagem para gerar uma imagem que represente o mundo descrito. Este passo permite que os usuários estabeleçam o cenário e a atmosfera para sua aventura de RPG.

Inicialização da História: Os usuários então fornecem uma descrição básica da história, delineando seu enredo e os elementos-chave. A aplicação utiliza Modelos de Linguagem para transformar essa entrada em uma narrativa mais detalhada, preparando o cenário para a aventura de TTRPG. Este passo marca a transição da preparação para o envolvimento ativo na história em evolução.

Tela Principal do Jogo: Agora que temos detalhes do personagem, do mundo e da história estabelecidos, os usuários acessam a tela principal do jogo. A aplicação envia informações relevantes para um Modelo de Linguagem, gerando dinamicamente um segmento narrativo juntamente com quatro escolhas distintas para o jogador. Os jogadores então fazem escolhas selecionando uma opção numerada, cada uma associada a um nível de dificuldade.

Tela de Rolagem de Dados: Após fazer uma escolha, os usuários avançam para a tela de “Rolagem de Dados”. Aqui, eles podem iniciar uma rolagem de dados, gerando um número aleatório entre 1 e 20. O resultado da rolagem do dado, combinado com o nível de dificuldade da opção escolhida, influencia o sucesso ou o fracasso da decisão do jogador.

Retorno à Tela Principal do Jogo: A aplicação envia a escolha do jogador, o resultado da rolagem do dado e o estado atual da história para o Modelo de Linguagem. O Modelo de Linguagem gera a próxima parte da história, incorporando as escolhas do usuário e a imprevisibilidade introduzida pela rolagem do dado. Este processo iterativo continua, permitindo uma narrativa em constante evolução moldada pelas decisões do usuário e elementos de chance.

No geral, o fluxo do usuário vai desde a criação de personagem até a geração dinâmica de história, criando uma experiência envolvente e personalizada de TTRPG.

6.2 DETALHES TÉCNICOS

Esta aplicação web de Simulação de RPG é construída usando “React”, uma biblioteca JavaScript para a criação de interfaces de usuário. A arquitetura é estruturada de modo a fornecer modularidade, manutenibilidade e um fluxo eficiente de dados entre os componentes.

Além do “React”, esta aplicação web também utiliza o “Vite” como uma ferramenta de construção. O “Vite” é uma ferramenta de construção front-end rápida e eficiente que aprimora a experiência de desenvolvimento ao fornecer compilações rápidas e otimizadas.

6.2.1 Estrutura de Componentes do React

StartScreen.jsx.jsx: A tela inicial onde os usuários apertam o botão "Start Game" para começar o jogo.

CreateCharacterScreen.jsx: A tela onde os usuários inserem o nome e a descrição do seu personagem.

Utiliza o componente de "estado" do React (useState) para gerenciar o nome e a descrição do personagem.

Realiza solicitações de API a um servidor de Modelo de Linguagem para gerar uma imagem com base na descrição do personagem.

WorldScreen.jsx: Permite que os usuários descrevam o cenário do mundo para a aventura de TTRPG.

Utiliza o componente de "estado" do React (useState) para gerenciar a descrição do mundo.

Realiza solicitações de API para gerar uma imagem que representa o mundo descrito.

CreateStoryScreen.jsx: Possibilita que os usuários definam o enredo básico e o cenário da história de TTRPG.

Utiliza o componente de "estado" do React (useState) para gerenciar o cenário da história.

Transiciona para a tela principal do jogo após a conclusão, passando informações essenciais para a próxima etapa.

GameScreen.jsx: O componente central onde a aventura principal de TTRPG se desenrola.

Utiliza os componentes (useState) e (useEffect) para gerenciar e atualizar vários estados, tais como o nome do personagem, as descrições do mundo e da história.

Realiza solicitações de API a um servidor de Modelo de Linguagem para gerar dinamicamente a narrativa com base nas escolhas do usuário.

DiceScreen.jsx: Apresenta a interface de rolagem de dados aos usuários após eles fazerem uma escolha no jogo.

Gerencia o estado da rolagem de dados (RolledFace) usando o componente (useEffect).

Permite que os usuários rolem os dados, desencadeando a geração de um número aleatório.

6.2.2 Solicitações de API

Geração de Personagens: Utiliza uma solicitação de API a um servidor de Modelo de Linguagem com um prompt específico para gerar uma imagem de personagem com base na descrição do personagem do usuário.

A resposta contém a URL da imagem gerada, que é então exibida como a imagem do personagem.

Geração de Mundos: Semelhante à geração de personagens, é feita uma solicitação de API a um servidor de Modelo de Linguagem com um prompt específico para gerar uma imagem de mundo.

A resposta contém a URL da imagem gerada, que é então exibida como a imagem do mundo.

6.3 EXPLICAÇÃO DOS COMPONENTES

CharacterScreen

Propósito: Interface para a criação de personagens.

Funcionalidade: Permite que o jogador insira o nome e a descrição do seu personagem. Envia uma solicitação a um LLM para gerar uma imagem com base nos detalhes fornecidos.

WorldScreen

Propósito: Interface para a criação do cenário do mundo.

Funcionalidade: Permite que o jogador descreva o mundo em que a história ocorre. Envia a descrição a um LLM, que retorna uma imagem representando o mundo descrito.

CreateStoryScreen

Propósito: Constrói a narrativa geral do TTRPG.

Funcionalidade: Fornece uma área de texto para o jogador inserir os elementos básicos da história do jogo, que serão posteriormente utilizados por um LLM gerador de texto para criar um texto detalhado com base no cenário fornecido.

A transição para a tela principal do jogo é habilitada após a conclusão.

GameScreen

Propósito: Serve como a interface principal para a simulação do TTRPG.

Funcionalidade: Exibe informações sobre personagens e o mundo, juntamente com o texto de história gerado dinamicamente. Gerencia o loop do jogo, onde o jogador toma decisões e o LLM gera eventos subsequentes.

Controla o fluxo entre eventos de história e escolhas do jogador.

DiceScreen

Propósito: Implementa o mecanismo de rolagem de dados.

Funcionalidade: Permite que o jogador role um dado de 20 lados ao clicar nele, produzindo assim um número aleatório. O resultado da rolagem é usado em conjunto com a escolha do jogador para determinar o sucesso ou fracasso da ação escolhida.

Possibilita a confirmação e progressão no loop do jogo.

Cada componente contribui para a simulação holística de TTRPG, permitindo que o jogador construa e experimente progressivamente uma narrativa única em um jogo de interpretação de papéis de mesa simulado.

6.4 INTERAÇÃO COM MODELOS DE LINGUAGEM

Na aplicação web de simulação de TTRPG, a interação com modelos de linguagem (LLMs) ocorre principalmente por meio de três processos-chave:

Geração de Personagem (`generateCharacterImage()`): Quando um jogador fornece o nome e a descrição do seu personagem na tela *CharacterScreen*, a aplicação utiliza a função `generateCharacterImage()` para fazer uma solicitação de API a um LLM. O LLM recebe um prompt contendo os detalhes do personagem (descrição do personagem) e gera uma imagem que representa o personagem. O LLM, então, retorna uma imagem com base nos detalhes do personagem fornecidos, a qual é exibida na aplicação.

Geração de Mundo (`generateWorldImage()`): Na tela *WorldScreen*, o jogador insere uma descrição do cenário do mundo para o TTRPG. A aplicação utiliza a função `generateWorldImage()` para enviar um prompt ao LLM contendo a descrição do mundo. O LLM, então, gera uma imagem que visualiza o mundo descrito e a envia de volta à aplicação para exibição.

Geração de História (*StoryEvents.js* para *GameScreen*): Durante o jogo na *GameScreen*, o jogador interage com a história gerada dinamicamente e faz escolhas. Após uma escolha do jogador, a aplicação utiliza as funções descritas em *StoryEvents.js* para escrever um prompt com base no texto da história em andamento, na ação do jogador e no valor do dado rolando. Esse prompt é então enviado ao LLM, que gera a próxima parte da história com base no contexto fornecido (texto anterior da história, ação do jogador e rolagem do dado). O texto gerado pelo LLM inclui possíveis desdobramentos e escolhas para o jogador considerar.

Prompts e Respostas com LLMs:

Prompts: Os prompts formulados incluem contexto do estado atual do jogo (descrições de personagem/mundo, história em andamento), ações do jogador e rolagens de dados aleatórias. Eles fornecem diretrizes específicas e contexto para que o LLM gere o próximo segmento relevante de texto.

Respostas: As respostas do LLM consistem em texto gerado dinamicamente, detalhando a progressão da história com base nas escolhas do jogador, rolagens de dados e no nível de dificuldade associado à ação escolhida. As respostas também incluem possíveis desfechos e escolhas subsequentes, estruturadas em um formato predefinido para manter a continuidade do jogo.

6.5 TELAS DA APLICAÇÃO

Em seguida, será mostrada uma sequência de telas equivalente a criar um mundo e jogar dois passos de um jogo.

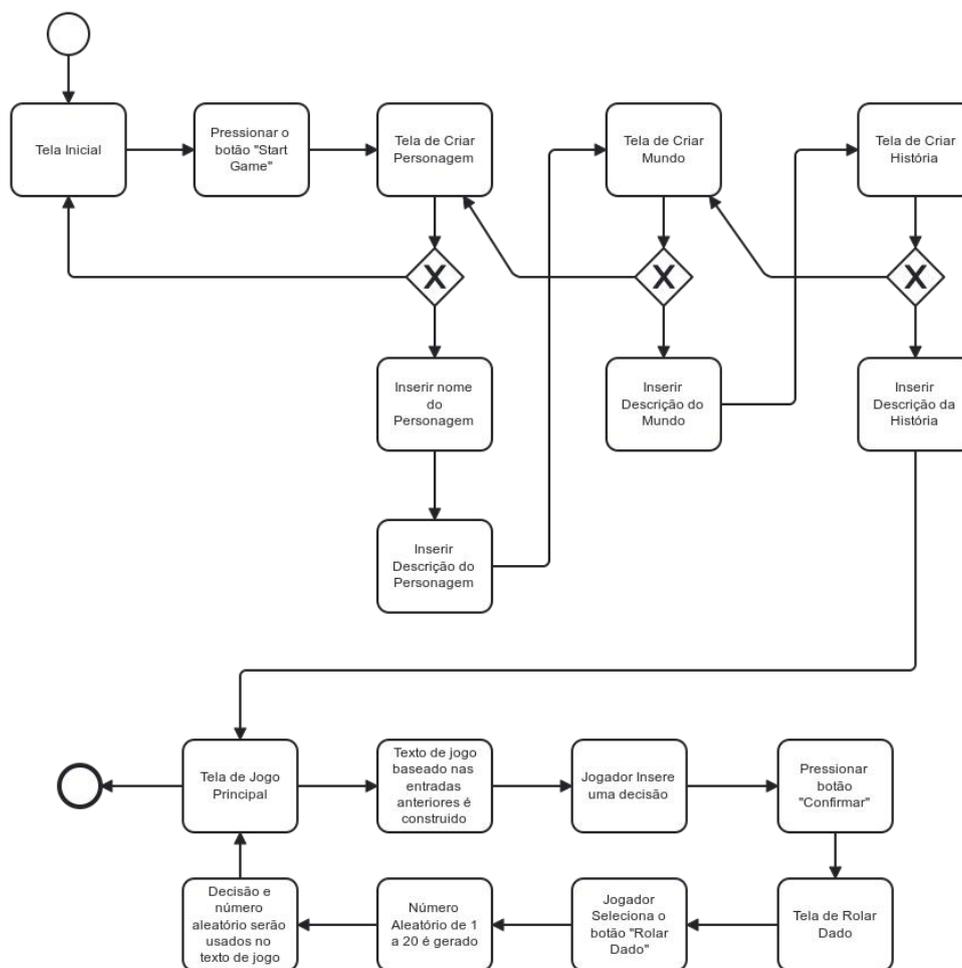


Figura 11 – Diagrama de Fluxo do Jogo

6.5.1 Telas Inicial

Tela Inicial da aplicação. O jogador pode começar o jogo ao apertar o botão "Start Game".

6.5.2 Criação de Personagem

Tela de criação de personagem. O jogador deve primeiro fornecer um nome para seu personagem e depois uma descrição. A aplicação então irá fazer uma solicitação de API para gerar uma imagem correspondente à descrição do personagem.

6.5.3 Criação do Mundo

Tela de criação do mundo. O jogador deve fornecer uma descrição para o mundo em qual o jogo ocorre. A aplicação então irá fazer uma solicitação de API para gerar uma imagem correspondente à descrição do mundo.



Figura 12 – Telas Inicial do Jogo

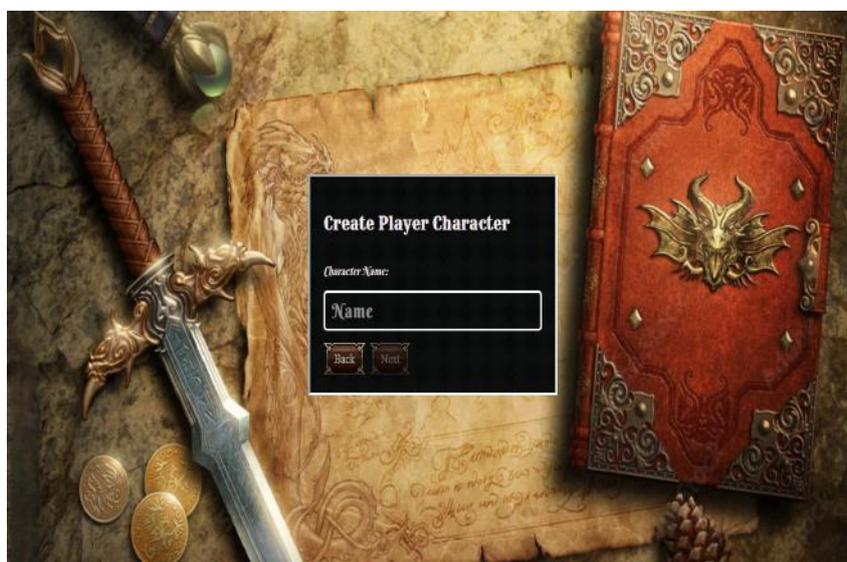


Figura 13 – Tela de Criação de Personagem

6.5.4 Criação da História

Tela de criação da história. O jogador deve fornecer o enredo básico da história, o qual será usado nas solicitação de API para gerar o texto da história.

6.5.5 Tela Principal do Jogo

Tela principal do jogo. A aplicação utiliza as entradas anteriores para o texto da história, gerando dinamicamente um segmento narrativo juntamente com quatro escolhas distintas para o jogador. Os jogadores então fazem escolhas selecionando uma opção numerada, cada uma associada a um nível de dificuldade. Após jogador mesmo informar



Figura 14 – Tela de Criação do Mundo

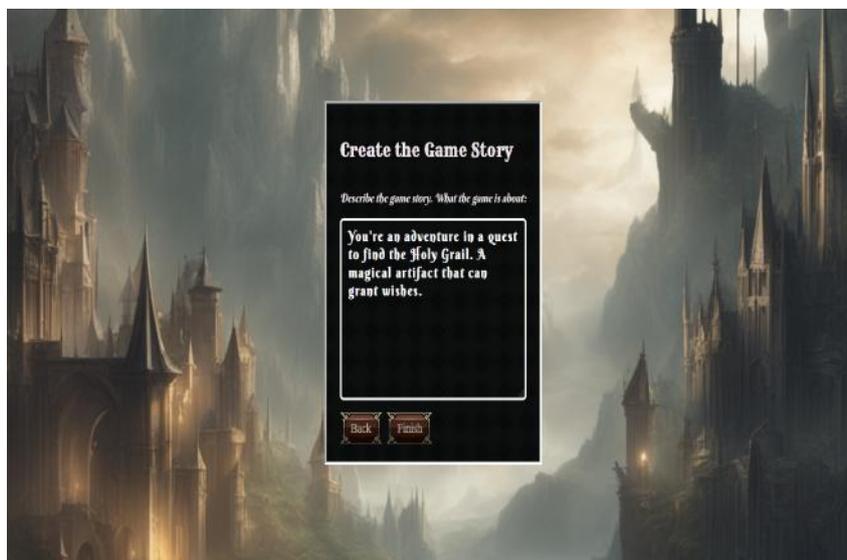


Figura 15 – Tela de Criação da História

sua escolha, o mesmo deve apertar o botão de confirmação.

6.5.6 Tela do Dado

Uma vez que o jogador tenha escolhido uma das opções oferecidas na tela de jogo, em seguida ele irá para a Tela do Dado. Aqui o jogador deve apertar o botão "Rolar Dado" e o mesmo irá gerar um número aleatório de 1 a 20. Ambas a escolha e o valor aleatório do dado serão mandado de volta para a Tela Principal do Jogo e usados para gerar o próximo segmento da história.

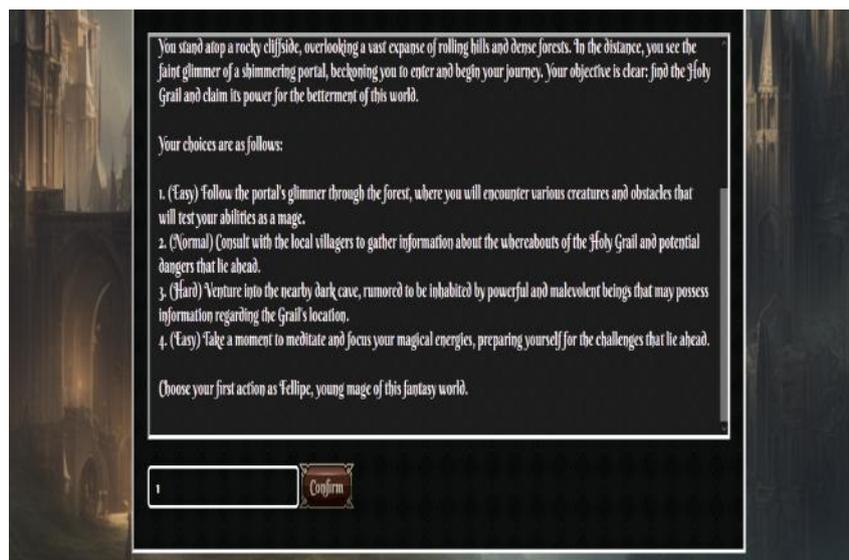


Figura 16 – Tela Principal do Jogo



Figura 17 – Tela do Dado

6.6 CONCLUSÃO

Esta aplicação web de simulação de TTRPG desenvolvida proporciona uma experiência imersiva e interativa de contar histórias para os jogadores. Começando pela criação de personagens, os usuários inserem detalhes sobre seus personagens, como nome e descrição, o que leva a aplicação a gerar uma imagem única do personagem usando um Modelo de Linguagem (LLM). Da mesma forma, o jogador descreve o cenário do mundo, e a aplicação utiliza outro LLM para criar uma imagem atmosférica correspondente.

Passando para a fase de criação de narrativa, os usuários definem o enredo básico, preparando o terreno para uma aventura envolvente. Uma vez tendo estabelecido os fundamentos, os jogadores entram na tela principal do jogo, onde suas escolhas são enviadas

a um LLM gerador de texto para criar a próxima parte da história. O texto gerado apresenta aos usuários quatro escolhas, cada uma associada a um nível de dificuldade. A escolha do jogador e uma rolagem de dados da tela de Rolagem de Dados ditam o sucesso ou fracasso, influenciando a direção da narrativa.

A arquitetura da aplicação depende de componentes “React”, estruturados para facilitar um fluxo de usuário sem interrupções, enquanto o framework “Vite” aprimora a velocidade de desenvolvimento. A aplicação se comunica com os LLMs para a geração de personagens e mundos por meio de solicitações de API, tais como `generateCharacterImage()` e `generateWorldImage()`, que utilizam prompts adaptados às descrições respectivas, extraíndo imagens únicas para personagens e mundos.

O módulo `StoryEvents.js` orquestra a interação com os LLMs durante o jogo. Ele lida tanto com a introdução inicial do jogo quanto com eventos de ação subsequentes desencadeados pelas escolhas do jogador. Essas interações seguem uma estrutura sistemática de prompts, garantindo uma narrativa coesa e envolvente. Os prompts incluem informações sobre o personagem, mundo, cenário e ações do jogador, criando um contexto abrangente para os LLMs.

Do ponto de vista técnico, o projeto envolve fazer chamadas de API assíncronas para os LLMs, gerenciar o estado por meio de componentes do “React” e lidar com a geração dinâmica de conteúdo. A inclusão de rolagens de dados adiciona um elemento estocástico, introduzindo incerteza, possibilidades e emoção à narrativa. O estilo temático, efeitos de animação e design responsivo contribuem para uma interface esteticamente agradável e amigável ao usuário.

Ao longo da implementação do TTRPG foram obtidas percepções (insights) valiosas sobre a integração de Modelos de Linguagem em aplicações web, o uso do “React” para o gerenciamento de estado e as complexidades de criar uma experiência envolvente e interativa de contar histórias. A combinação de elementos narrativos criativos, entrada do usuário e geração dinâmica de conteúdo torna esta simulação de TTRPG uma aplicação web inovadora e divertida.

7 CONCLUSÃO

Ao longo deste trabalho sobre a interseção entre Inteligência Artificial (IA) e videogames, examinamos profundamente como a aplicação da IA, especialmente a "IA Generativa" e os "Modelos de Linguagem Grandes" (LLMs), tem o potencial de redefinir o campo de videogames. Com uma análise detalhada e uma abordagem multifacetada, este trabalho busca resumir as conclusões obtidas e as implicações dessa pesquisa.

O objetivo central foi compreender não apenas o estado atual dessa interseção, mas também antever as possibilidades futuras e os desafios a serem enfrentados. Neste capítulo final, consolidaremos as principais conclusões retiradas deste estudo e ofereceremos insights sobre o impacto e o caminho futuro dessa dinâmica relação.

Ao revisarmos a história da IA em videogames, testemunhamos uma evolução notável desde as formas rudimentares até as aplicações mais sofisticadas de hoje. A IA deixou de ser meramente uma ferramenta reativa, tornando-se proativa, adaptativa e capaz de criar experiências mais ricas e desafiadoras para os jogadores.

A análise do funcionamento dos Chatbots de IA e LLMs revelou a profundidade e a complexidade dessas tecnologias. Os Chatbots, alimentados por modelos como GPT-3, mostraram-se capazes de interações naturais, enquanto os LLMs demonstraram uma notável capacidade de gerar texto coerente e contextualmente relevante.

A análise da "IA Generativa" destacou seu papel inovador na criação de narrativas dinâmicas. A capacidade de gerar conteúdo narrativo em tempo real promete redefinir a forma como os jogadores interagem com as histórias dos jogos, oferecendo experiências mais personalizadas e envolventes.

A exploração das aplicações práticas da integração de LLMs e "IA Generativa" em jogos revelou uma série de exemplos interessantes. Desde a personalização de diálogos até a geração de ambientes dinâmicos, essas tecnologias estão se tornando peças fundamentais no arsenal dos desenvolvedores de jogos.

Por último, a implementação prática do Simulador de TTRPG forneceu uma perspectiva valiosa sobre como as teorias discutidas podem ser aplicadas na criação de uma experiência de jogo interativa e única. Essa aplicação permitiu que os jogadores moldassem ativamente suas narrativas, empregando uma interseção de escolhas do jogador e eventos gerados pela IA.

7.1 CONSIDERAÇÕES FINAIS

Em resumo, este estudo proporcionou uma imersão abrangente na interseção entre Inteligência Artificial e videogames. A rápida evolução dessas tecnologias promete não

apenas transformar a forma como jogamos, mas também desafiar as fronteiras da criatividade e da personalização.

Para o futuro, é imperativo que os desenvolvedores e pesquisadores continuem a equilibrar essas inovações com as devidas considerações éticas. A busca por experiências de jogo mais dinâmicas e envolventes deve ser acompanhada por uma reflexão constante sobre os impactos sociais e culturais dessas tecnologias.

Por fim, percebemos que a simbiose entre IA e videogames está longe de atingir seu auge. Devemos então permanecer atentos às futuras possibilidades e aos desafios instigantes que aguardam nessa integração fascinante da tecnologia e da diversão interativa.

REFERÊNCIAS

- ADAMOPOULOU, E. An overview of chatbot technology. **Artificial Intelligence Review**, 2020. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7256567/>.
- AUERBACH, D. The programs that become the programmers. **Slate**, 2015. Disponível em: <https://slate.com/technology/2015/09/pedro-domingos-master-algorithm-how-machine-learning-is-reshaping-how-we-live.html>.
- BANH, L.; STROBEL, G. Generative artificial intelligence. **Electronic Markets**, 2023.
- BEDINGFIELD, W. Generative ai won't revolutionize game development just yet. **Wired UK**, 2023. Disponível em: <https://www.wired.co.uk/article/generative-ai-video-game-development>.
- BOOTH, M. **The AI Systems of Left 4 Dead**. 2009. Disponível em: https://steamcdn-a.akamaihd.net/apps/valve/2009/ai_systems_of_l4d_mike_booth.pdf.
- BROWN, T. B. et al. Language models are few-shot learners. In: **ACM. Proceedings of the 34th International Conference on Neural Information Processing Systems**. 2020. Disponível em: <https://dl.acm.org/doi/abs/10.5555/3495724.3495883>.
- BRUNDAGE, M. e. a. Prepare for truly useful large language models. **Nature Biomedical Engineering**, 2023. Disponível em: <https://www.nature.com/articles/s41551-023-01012-6>.
- CHEIGH, J. Generating images using vaes, gans, and diffusion models. **Towards Data Science**, 2023. Disponível em: <https://towardsdatascience.com/generating-images-using-vaes-gans-and-diffusion-models-48963ddeb2b2>.
- CLARK, J.; PERRAULT, R. **Artificial Intelligence Index Report 2022**. 2022. Disponível em: https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf.
- CRETU, C. How does chatgpt actually work? an ml engineer explains. **Scalable Path**, 2023. Disponível em: <https://www.scalablepath.com/machine-learning/chatgpt-architecture-explained>.
- DEVELOPERS, G. **The Impact of AI on Game Design Jobs**. 2023. Disponível em: <https://medium.com/game-developers/the-impact-of-ai-on-game-design-jobs-90b19dda0d0a>.
- EAPEN, T. T. et al. How generative ai can augment human creativity. **Harvard Business Review**, 2023. Disponível em: <https://hbr.org/2023/07/how-generative-ai-can-augment-human-creativity>.
- EDUCATION), E. E. at C. T. Unleashing the potential of large language models (llms) with chatgpt. **Medium**, 2023. Disponível em: <https://medium.com/@eugene-s/unleashing-the-potential-of-large-language-models-llms-with-chatgpt-8210f0cb063d>.

- EL-NASR, M. S. Façade: Experiencing interactive narrative: A qualitative analysis of façade. **Entertainment Computing**, 2013. Disponível em: https://www.researchgate.net/publication/257708815_Experiencing-interactive-narrative-A-qualitative-analysis-of-Facade.
- FERRARA, E. Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. **arXiv preprint arXiv:2304.07683**, 2023.
- GAMES, R. B. **Pathfinding Basics**. 2014. Disponível em: <https://www.redblobgames.com/pathfinding/a-star/introduction.html>.
- GARG, R. Training on large datasets that don't fit in memory in keras. **Medium**, 2019. Disponível em: <https://medium.com/@mrgarg.rajat/training-on-large-datasets-that-dont-fit-in-memory-in-keras-60a974785d71>.
- GOODFELLOW, I. J. et al. Generative adversarial nets. **arXiv preprint arXiv:1406.2661**, 2014. Disponível em: <https://arxiv.org/pdf/1406.2661.pdf>.
- GREEN, M.; KHALIFA, A. Generating levels that teach mechanics. In: **Proceedings of the [Full Conference Name]**. [s.n.], 2018. Disponível em: https://www.researchgate.net/publication/327640362_Generating_levels_that_teach_mechanics.
- GRIMMELMANN, J.; LEE, K.; COOPER, A. F. Talkin' 'bout ai generation: Copyright and the generative ai supply chain. In: . [s.n.], 2023. Disponível em: <https://james.grimmelman.net/files/articles/talkin-bout-ai-generation.pdf>.
- GWERTZMAN, J.; SOSLOW, J. The generative ai revolution in games. **Andreessen Horowitz**, 2022. Disponível em: <https://a16z.com/the-generative-ai-revolution-in-games/>.
- HASTINGS, E. J.; GUHA, R. K.; STANLEY, K. O. Evolving content in the galactic arms race video game. In: . [s.n.], 2009. Disponível em: https://www.researchgate.net/publication/221157637_Evolving_content_in_the_Galactic_Arms_Race_video_game.
- HENDRIKX, M.; MEIJER, S.; VELDEN, J. V. D. Procedural content generation for games: A survey. **ACM Transactions on Multimedia Computing, Communications and Applications**, 2013. Disponível em: https://www.researchgate.net/publication/262327212_Procedural-Content-Generation-for-Games-A-Survey.
- HU, W.; ZHANG, Q.; MAO, Y. Component-based hierarchical state machine – a reusable and flexible game ai technology. **IEEE Conference Publication**, 2011. Disponível em: https://www.researchgate.net/publication/252047555_Component-based_hierarchical_state_machine_-_A_reusable_and_flexible_game_AI_technology.
- ISLA, D. Handling complexity in the halo 2 ai. In: **Game Developers Conference**. [s.n.], 2005. Disponível em: <https://www.gamedeveloper.com/programming/gdc-2005-proceeding-handling-complexity-in-the-i-halo-2-i-ai>.
- IVAN, T. Ubisoft is developing an ai tool 'that aims to support scriptwriters'. **Video Games Chronicle**, 2023. Disponível em: <https://www.videogameschronicle.com/news/ubisoft-is-developing-an-ai-tool-that-aims-to-support-scriptwriters/>.

- LAI, J. The neverending game: How ai will create a new category of games. **Andreessen Horowitz**, 2023. Disponível em: <https://a16z.com/the-neverending-game-how-ai-will-create-a-new-category-of-games/>.
- LIN, Y. et al. Pre-trained models for representation learning. In: _____. **Representation Learning for Natural Language Processing**. [s.n.], 2023. p. 127–167. Disponível em: https://www.researchgate.net/publication/373358021_Pre-trained_Models_for_Representation_Learning.
- LOCK, S. What is ai chatbot phenomenon chatgpt and could it replace humans? **The Guardian**, 2022. Disponível em: <https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans>.
- MAGUIRE, G. The astounding power of large language models (llms) in 2023. **Matrix Consulting Blog**, 2023. Disponível em: <https://www.matrixconsulting.ai/blog/a-guide-to-large-language-models-in-2023>.
- MARTIN, J. Angelina: The computer program that designs games. **Game Developer**, 2013. Disponível em: <https://www.gamedeveloper.com/design/angelina-the-computer-program-that-designs-games>.
- MARTINEZ, J. J. **AI in video games: a historical evolution, from Search Trees to LLMs. Chapter 1: 1950–1980**. 2023.
- MCGUIRE, T. Attention, transformers, and gpt. **Trevor McGuire Medium**, 2023. Disponível em: <https://trevormcguire.medium.com/attention-transformers-and-gpt-b3adbbb4a950>.
- MORIARTY, C. E. **The Legal Challenges of Generative AI—Part 1 Skynet and HAL Walk Into a Courtroom**. 2023. Disponível em: <https://cl.cobar.org/features/the-legal-challenges-of-generative-ai-part-1/>.
- NEWPORT, C. What kind of mind does chatgpt have? **The New Yorker**, 2023. Disponível em: <https://www.newyorker.com/science/annals-of-artificial-intelligence/what-kind-of-mind-does-chatgpt-have>.
- OSBOURN, T. **Space Invaders and the fun reason it gets progressively harder**. 2023. Written by Toby Osbourn. Disponível em: <https://tosbourn.com/space-invaders/>.
- PORTER, B. **Dynamic vs Static Procedural Generation**. 2017. Disponível em: <https://medium.com/@eigenbom/dynamic-vs-static-procedural-generation-ed3e7a7a68a3>.
- RABBIT, R. **AI In Games: Complicated Characters**. 2023. Disponível em: <https://medium.com/curiouserinstitute/ai-in-games-complicated-characters-6c47797d997a>.
- RISI, S.; TOGELIUS, J. Neuroevolution in games: State of the art and open challenges. **IEEE Transactions on Computational Intelligence and AI in Games**, IEEE, 2017.
- SANTIAGO, J. M. et al. Rolling the dice: Imagining generative ai as a dungeons & dragons storytelling companion. **arXiv preprint arXiv:2304.01860**, 2023.

SCHELL, J. **The Future of Storytelling: How Medium Shapes Story**. 2013. Disponível em: <https://www.gdcvault.com/play/1018026/The-Future-of-Storytelling-How>.

SIMPSON, C. Behavior trees for ai: How they work. **Game Developer**, 2014. Disponível em: <https://www.gamedeveloper.com/programming/behavior-trees-for-ai-how-they-work>.

SULLIVAN, M. **Why everyone seems to disagree on how to define Artificial General Intelligence**. 2023. Disponível em: <https://www.fastcompany.com/90968623/why-everyone-seems-to-disagree-on-how-to-define-artificial-general-intelligence>.

TEAM, B. E. **Small Generative AIs: Size Does Matter**. 2023. Disponível em: <https://www.bairesdev.com/blog/small-generative-ais-size-does-matter/>.

TEAM, D. A. **Scaling Generative AI Models: Challenges and Solutions**. 2023. Disponível em: <https://deci.ai/blog/scaling-generative-ai-models/>.

THOMAS, S. **Cloud Computing Is the Engine Powering Generative AI**. 2023. Disponível em: <https://individuals.voya.com/insights/investment-insights/cloud-computing-engine-powering-generative-ai>.

TOGELIUS, J. What is procedural content generation? mario on the borderline. In: **Proceedings of the 4th International Conference on Computational Creativity**. [s.n.], 2011. Disponível em: https://www.researchgate.net/publication/228620622_What-is-Procedural-Content-Generation-Mario-on-the-borderline.

TOGELIUS, J. **Playing Smart: On Games, Intelligence, and Artificial Intelligence**. [s.n.], 2019. Disponível em: <https://direct.mit.edu/books/book/4160/Playing-SmartOn-Games-Intelligence-and-Artificial>.

V, V. How chatgpt overcomes the ai alignment problem? **Medium**, 2023. Disponível em: <https://medium.com/@vignesh4757/how-chatgpt-overcomes-the-ai-alignment-problem-df7d0f348dd9>.

VASWANI, A. e. a. **Transformer: A Novel Neural Network Architecture for Language Understanding**. Google Research, 2017. Posted by Jakob Uszkoreit, Software Engineer, Natural Language Understanding. Disponível em: <https://blog.research.google/2017/08/transformer-novel-neural-network.html>.

WEI, J. et al. **Emergent Abilities of Large Language Models**. 2022. Disponível em: <https://arxiv.org/abs/2206.07682>.

WINTERMEYER, L. **Generative AI Is Generating Hype But Can It Generate Value?** 2023. Disponível em: <https://www.forbes.com/sites/lawrencewintermeyer/2023/10/26/generative-ai-is-generating-hype-but-can-it-generate-value/?sh=26313ddae81f>.

WOLFRAM, S. What is chatgpt doing ... and why does it work? **Stephen Wolfram Writings**, 2023. Disponível em: <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>.

WRIGHT, S. T. Blizzard is reportedly set to unveil its own ai tool for concept art. **GameSpot**, 2023. Disponível em: <https://www.gamespot.com/articles/blizzard-is-reportedly-set-to-unveil-its-own-ai-tool-for-concept-art/1100-6514464/>.

WU, Z. et al. Ai creativity and the human-ai co-creation model. In: KUROSU, M. (Ed.). **Human-Computer Interaction. Theory, Methods and Tools**. Cham: Springer International Publishing, 2021. p. 171–190. ISBN 978-3-030-78462-1.

YANNAKAKIS, G. N.; TOGELIUS, J. **Artificial Intelligence and Games**. 2018. Disponível em: <https://gameaibook.org/book.pdf>.

ZVORNICANIN, E. How does chatgpt work? **Baeldung**, 2023. Disponível em: <https://www.baeldung.com/cs/chatgpt-model>.