

**“Do conhecimento acadêmico à transformação sustentável: inovação com validação científica”****ESTUDO DE IA E APLICAÇÃO EM APRENDIZAGEM POR REFORÇO PARA JOGOS**Letícia Borges dos Reis<sup>1</sup> (EG), Felipe Bastos Vargas<sup>1</sup> (EG), Egon Luiz Muller<sup>1</sup> (PQ)<sup>1</sup>Universidade Federal de Itajubá- Campus Itajubá.**Palavras-chave:** Inteligência Artificial. Aprendizagem por Reforço. Q-Learning; Jogos Digitais. OpenAI Gym.**Introdução**

O avanço da Inteligência Artificial tem destacado a necessidade de compreender não apenas os algoritmos de aprendizado empregados, mas também a forma como são implementados, uma vez que tais escolhas impactam diretamente o desempenho alcançado. Portanto, mais do que a simples aplicação de técnicas de aprendizado, ressalta-se a importância de garantir implementações eficientes, capazes de produzir resultados consistentes e de elevado desempenho.

Por outro lado, no âmbito da Aprendizagem por Reforço, o método de exploração adotado exerce papel fundamental, uma vez que regula o equilíbrio entre a busca por novas possibilidades e o aproveitamento do conhecimento previamente adquirido. Essa escolha pode determinar se o agente alcançará uma solução ótima de forma rápida ou se permanecerá restrito a estratégias limitadas.

Nesse sentido, o presente projeto tem como propósito analisar o impacto de diferentes estratégias de exploração no desempenho de agentes em um ambiente de jogo. Para tanto, emprega-se o algoritmo Q-Learning, que possibilita ao agente aprender por meio da interação com o ambiente, ajustando progressivamente suas decisões até a convergência para políticas mais eficazes.

Diversos métodos de exploração foram implementados, cada um com características específicas:  $\epsilon$ -greedy, que alterna entre escolhas aleatórias e ações conhecidas; Boltzmann Exploration (Softmax), que atribui probabilidades às ações a partir dos valores de Q; Greedy with Random Tie-Breaking, que resolve empates de forma aleatória; Exploração Dirigida por Curiosidade (RND simplificado), que incentiva a visita a estados pouco explorados; e Reward Shaping para Exploração, que introduz recompensas temporárias adicionais a fim de estimular novas trajetórias.

A avaliação foi conduzida em três mapas distintos de 8x8, contemplando diferentes formas de pontuação:

menor número de passos, maior quantidade de moedas coletadas ou maior pontuação total (com perda de um ponto por passo e ganho de três por moeda). Também foi analisado o comportamento dos agentes em cenários com o recurso slippery ativado, no qual há incerteza nos movimentos, tornando o aprendizado mais desafiador.

Por fim, o desempenho foi mensurado segundo três parâmetros principais: tempo necessário para atingir o melhor caso, pontuação obtida em cada episódio e número de tentativas até alcançar o resultado desejado.

**Metodologia**

O início do desenvolvimento do projeto teve como base o uso da linguagem python e o uso de bibliotecas com jogos integrados, como o Gym e Gymnasium, que facilitam a realização da análise voltada para o desempenho da inteligência artificial e suas diferentes aplicações e variações de métodos de exploração, como mencionado anteriormente.

Tendo definidas as ferramentas a serem utilizadas, foi implementado o ambiente do jogo Frozen Lake utilizando da biblioteca Gym, da OpenAI. Em seguida, o algoritmo Q-Learning foi aplicado com base nos movimentos do personagem ao longo do mapa pré definido, onde cada movimento teria um valor atribuído de acordo com sua eficiência em chegar ao objetivo final da fase no jogo.



Figura 1 – Ambiente do jogo Frozen Lake

## “Do conhecimento acadêmico à transformação sustentável: inovação com validação científica”

Após a primeira etapa, o desenvolvimento de um ambiente próprio se mostrou mais efetivo para dar prosseguimento, permitindo que alterações fossem feitas na base de como o jogo funciona para que se enquadre melhor com as necessidades do projeto.

Agora com um ambiente funcional e personalizável, e a estratégia base de Q-Learning implementada, foram exploradas as possibilidades de diferentes funcionamentos do algoritmo e quais seriam os mais adequados a serem implementados como os métodos de exploração que seriam comparados de acordo com os parâmetros definidos de pontuação desejada.

Sobre a implementação das diferentes pontuações, basta mudar os parâmetros em que os valores atribuídos do Q-learning se baseia, onde para o modo base, em que o objetivo é apenas chegar no final da fase sem falhar, são atribuídos pontos positivos para cada movimento que leva o personagem para mais perto do objetivo final, e na menor quantidade de passos. Para o treinamento com pontuação por moedas, basta trocar o que antes seria a proximidade do fim da fase pela coleta de moedas, onde ações que fazem com que o personagem se torne mais provável de coletar uma moeda é o cenário ideal. Para maior pontuação total, a tabela de valores da IA é baseada tanto na distância do final, quantidade de passos e das moedas coletadas.

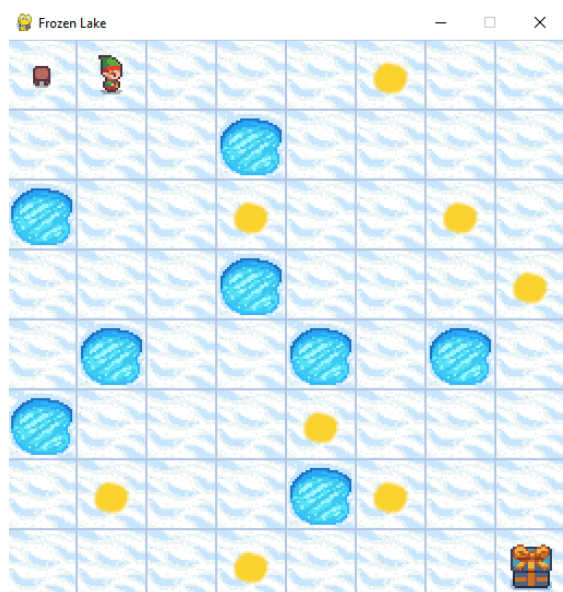


Figura 2 – Mapa personalizado e moedas implementadas

Até então, esta é a última parte já implementada do projeto faltando agora coletar dados de treinamento da IA ao longo de diversas tentativas, e explorando as diferentes combinações de método de exploração e formas de pontuação, e após a coleta dos dados de número de tentativas, pontuação obtida, e o tempo de treinamento do algoritmo em si. Esses dados serão tabelados para se ter uma melhor perspectiva em qual situação o uso de cada método de exploração se mostra o ideal.

### Resultados e discussão

Embora o objetivo final ainda não tenha sido alcançado no momento da escrita deste documento, a principal parte da pesquisa já está operacional e se mostrando promissora para a aplicação final.

Um possível problema para o treinamento em larga escala das IAs vem do gasto computacional e tempo necessário para que sejam completos sem erros, uma vez que para o teste dos métodos de exploração e se estão agindo corretamente, não é necessário uma grande base de iterações, mas para a coleta de dados, se torna mais notável.

Outra discussão abordada é sobre quantos episódios e profundidade da exploração são necessários para se ter um resultado satisfatório, sem ter que, novamente, sacrificar muito uso computacional.

Como o projeto ainda não está encerrado, caso se mostra necessário, o estudo do número de episódios e parâmetros internos dos métodos de exploração também podem ser medidos para a entrega final, mostrando a eficiência de cada caso e em que momento se torna mais custoso do que benéfico os diferentes parâmetros definidos para a exploração.

### Conclusões

Conclui-se que a Inteligência Artificial tende a estar cada vez mais presente em diferentes áreas da sociedade, ampliando seu impacto e relevância. Nesse cenário, torna-se essencial que sua implementação seja realizada de forma eficiente e condizente com os objetivos propostos, garantindo não apenas melhores resultados, mas também a adequação de sua aplicação aos contextos em que é empregada. Considerando que a

## “Do conhecimento acadêmico à transformação sustentável: inovação com validação científica”

análise de desempenho ainda será desenvolvida, espera-se que os resultados obtidos permitam verificar de forma clara como os diferentes métodos de exploração influenciam o desempenho do agente, contribuindo para a compreensão de sua aplicação em ambientes de jogos e, por extensão, em outras áreas práticas da Inteligência Artificial.

### Agradecimentos

Gostaríamos de expressar nossos agradecimentos à Universidade Federal de Itajubá e ao PET-TEC, por proporcionar os recursos necessários que possibilitaram a realização deste projeto. Além disso, estendemos nossa gratidão ao FNDE, por todo apoio.

### Referências

SUTTON, Richard S.; BARTO, Andrew G. **Reinforcement Learning: An Introduction**. 2. ed. Cambridge: MIT Press, 2018.

BRAGA, A. de P.; CARVALHO, A. C. P. de L. F.; LUDERMIR, T. B. **Redes Neurais Artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2007.

OPENAI. **Gym Documentation**. Disponível em: <https://www.gymnasium.dev/>. Acesso em: 10 jun. 2025.

PROGRAMAÇÃO DINÂMICA. **Aprendizado por Reforço (Reinforcement Learning)**. YouTube, 2020. Disponível em: <https://www.youtube.com/watch?v=zmYhF8gk5wE>. Acesso em: 17 abr. 2025.