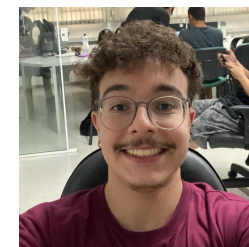



# Reinforcement Learning

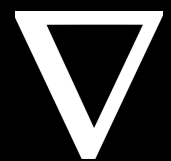


João Pedro Godoy  
 @joaopgodoy



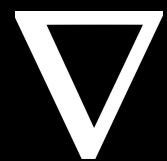
# Presença

- Linktree: Presente na bio do nosso instagram
- Presença ficará disponível até 1 hora antes da próxima aula
- É necessário 70% de presença para obter o certificado

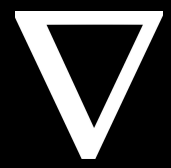


# Presença



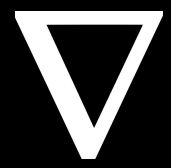
The background features a large, stylized 'V' shape composed of three overlapping triangles in light blue, light pink, and light purple. A blue L-shaped line is on the left, and a pink L-shaped line is on the right, both framing the central text.

# Introdução



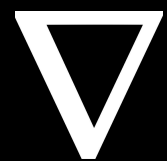
# O que é Inteligência Artificial?

É o campo da ciência da computação que desenvolve sistemas capazes de realizar tarefas que normalmente seriam realizadas por meio da inteligência humana.

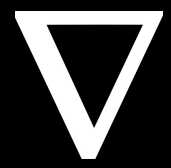


# O que é Inteligência Artificial?

- Tomar decisões
- Reconhecer padrões
- **Adaptar-se ao ambiente e aprender com a experiência**

The background features a large, light blue inverted triangle and a slightly offset, semi-transparent pink inverted triangle. A blue horizontal line and a pink horizontal line extend from the left and right sides of the text, respectively, meeting at the top and bottom of the text area.

# **Motivações e problemas**



# Algumas motivações da IA

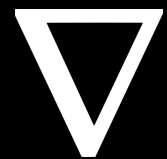
- Automatizar tarefas repetitivas
- Auxiliar em tarefas humanas como diagnóstico médico
- Melhorar eficiência e tomada de decisões em diversos setores
- Facilitar interações mais naturais entre humanos e máquina (linguagem natural)



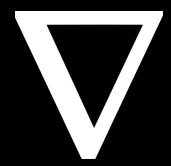


# Problemas que a IA pode resolver

- Reconhecimento e tradução de fala
- Classificação de imagens
- Desenvolvimento de agentes autônomos (carros, robôs)
- Sistemas de recomendação (filmes, músicas, compras online)
- Geração de textos, imagens, vídeos, música
- Detecção de fraudes bancárias

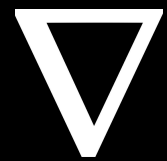
The background features a large, stylized graphic consisting of two overlapping triangles. The left triangle is light blue and the right triangle is light pink. They overlap in the center, creating a purple hue. A blue line starts from the left edge, extends horizontally, and then turns vertically downwards. A pink line starts from the right edge, extends horizontally, and then turns vertically upwards. The word "Paradigmas" is centered between these two lines.

# Paradigmas

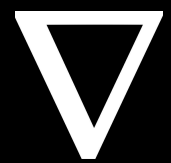


# Paradigmas de Aprendizado em IA

- Aprendizado Supervisionado
- Aprendizado Não Supervisionado
- **Aprendizado por Reforço (RL)**

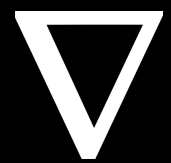
The background features a large, light blue downward-pointing triangle and a slightly offset, semi-transparent pink downward-pointing triangle. A blue horizontal line extends from the left edge of the text area, and a pink horizontal line extends from the right edge. Both lines have short vertical segments at their ends, creating a bracket-like effect around the text.

# **Aprendizado por Reforço**



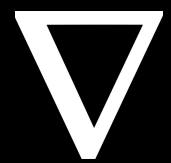
# O que é Aprendizado por Reforço?

Abordagem em que o sistema aprende por meio de **interação com o ambiente**, ajustando-se conforme o recebimento de **recompensas** ou **punições**.



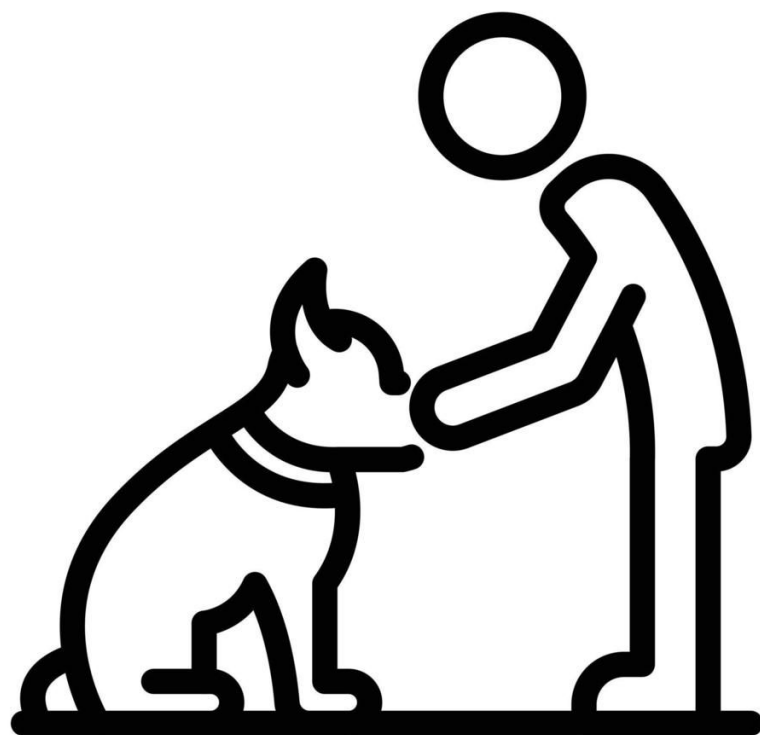
# O que é Aprendizado por Reforço?

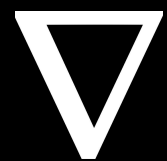
*Reinforcement Learning é aprender a tomar decisões com base na experiência, explorando ações para maximizar recompensas no longo prazo.*



# Exemplos intuitivos

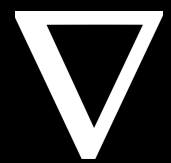
- Bebê aprendendo a andar
- Cachorro aprendendo a dar a pata



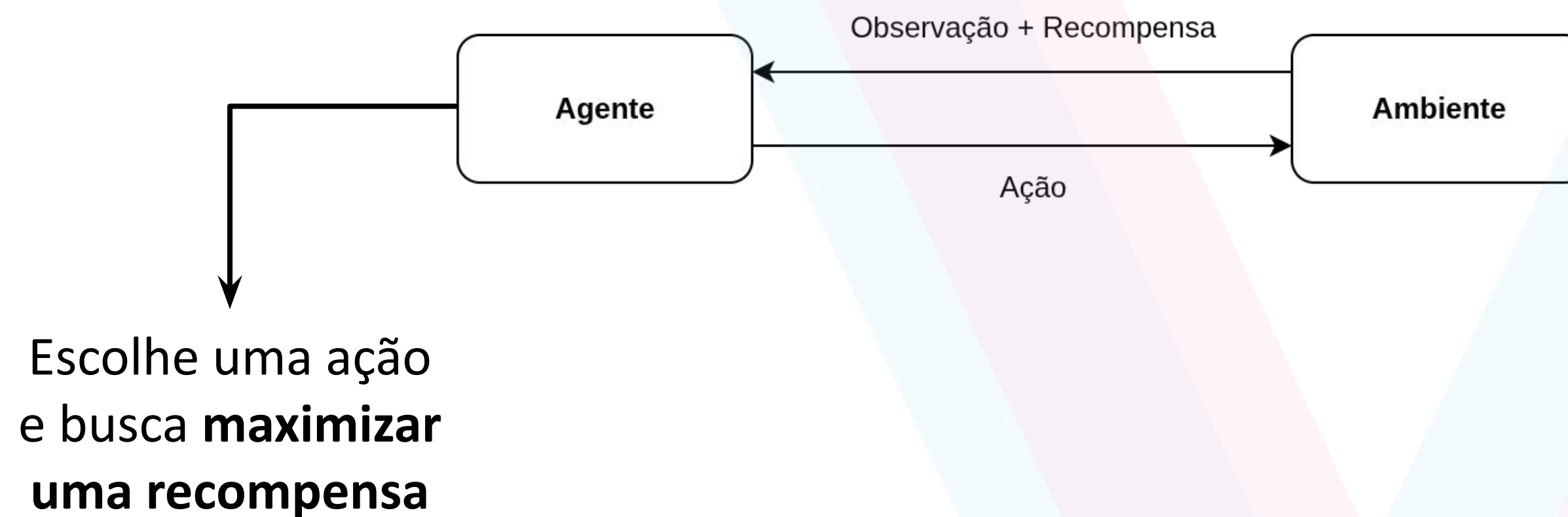
The background features a large, light blue inverted triangle and a slightly offset, semi-transparent pink inverted triangle. A light blue L-shaped line is on the left, and a pink L-shaped line is on the right, framing the central text.

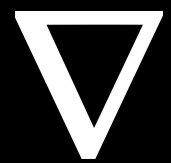
# **Elementos do Aprendizado por Reforço**



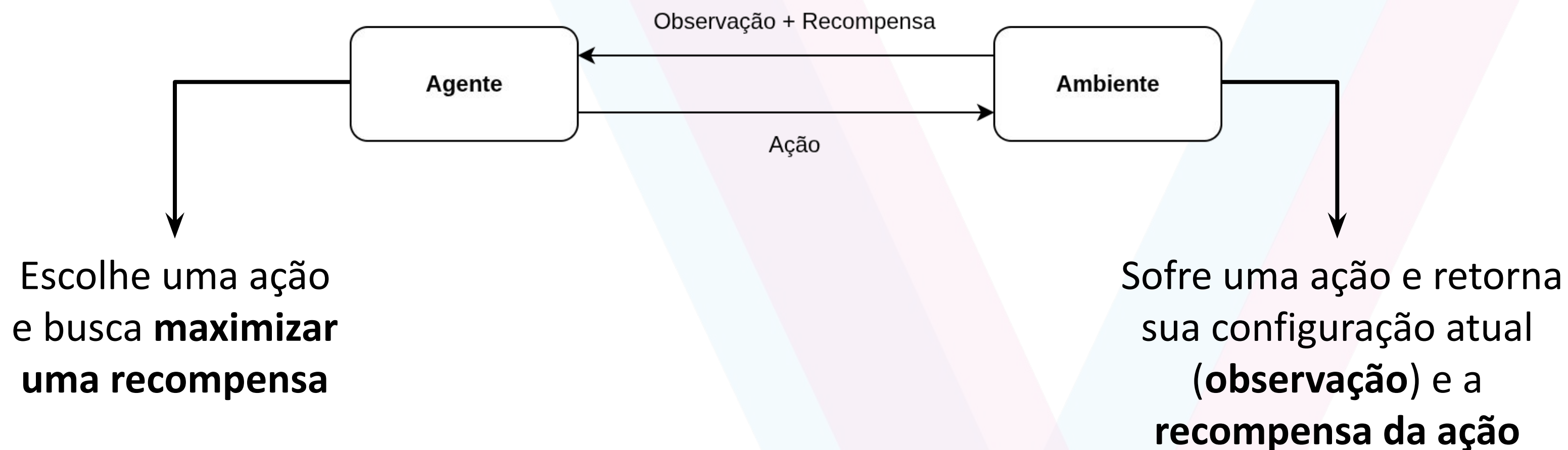


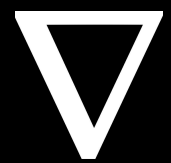
# Loop Ambiente-Agente





# Loop Ambiente-Agente

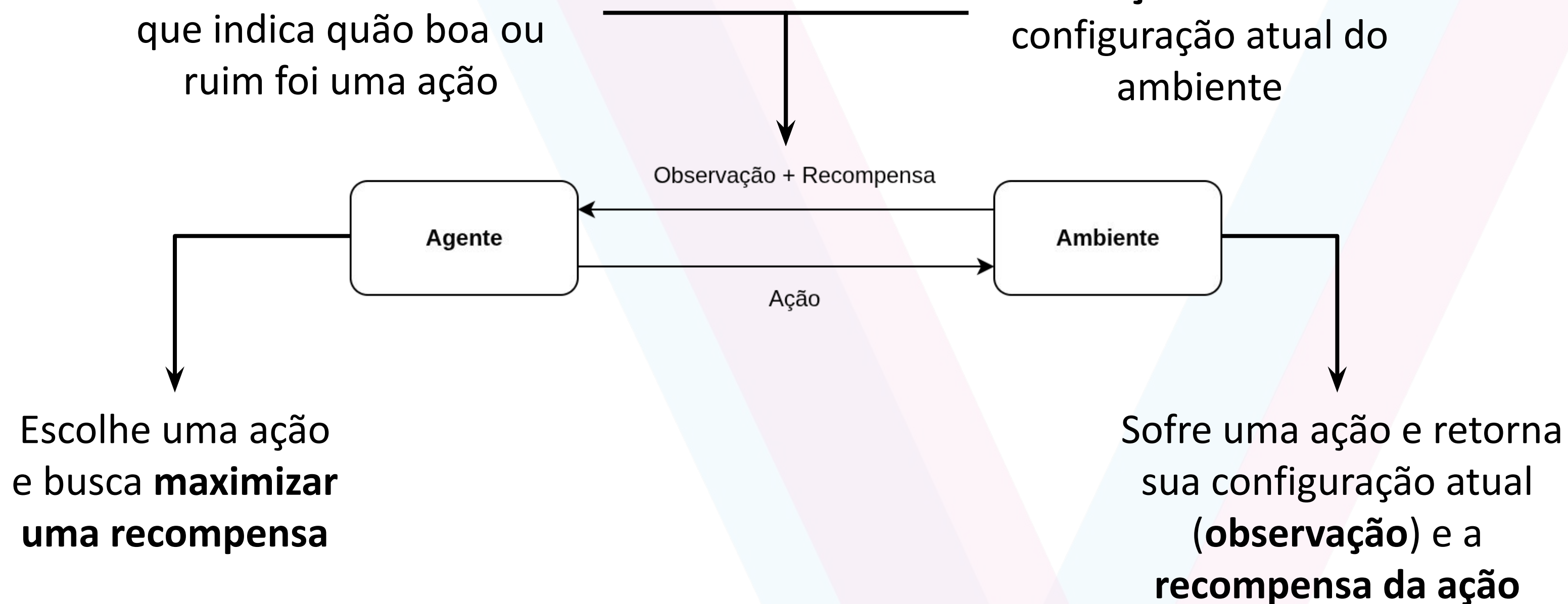


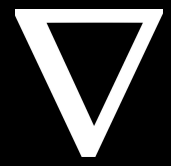


# Loop Ambiente-Agente

**Recompensa:** Número que indica quão boa ou ruim foi uma ação

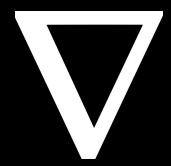
**Observação:** variáveis de configuração atual do ambiente





# Exemplo - Robô

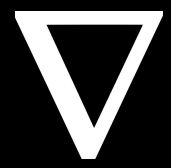
Imagine um robô que se move por um labirinto em busca de uma saída



# Exemplo - Robô

Imagine um robô que se move por um labirinto em busca de uma saída

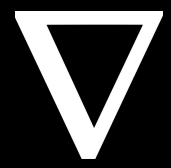
- **Agente:** Robô



# Exemplo - Robô

Imagine um robô que se move por um labirinto em busca de uma saída

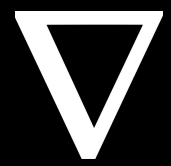
- **Agente:** Robô
- **Ambiente:** labirinto com obstáculos e saída



# Exemplo - Robô

Imagine um robô que se move por um labirinto em busca de uma saída

- **Agente:** Robô
- **Ambiente:** labirinto com obstáculos e saída
- **Ação:** qual o movimento do robô



# Exemplo - Robô

Imagine um robô que se move por um labirinto em busca de uma saída

- **Agente:** Robô
- **Ambiente:** labirinto com obstáculos e saída
- **Ação:** qual o movimento do robô
- **Observação:** posição do robô e dos obstáculos próximos





# Exemplo - Robô

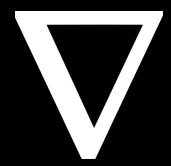
Imagine um robô que se move por um labirinto em busca de uma saída

- **Agente:** Robô
- **Ambiente:** labirinto com obstáculos e saída
- **Ação:** qual o movimento do robô
- **Observação:** posição do robô e dos obstáculos próximos
- **Recompensa:** -1 se bateu, 0 se andou sem batida e +10 se atingiu a saída



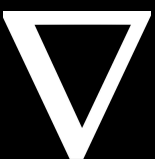
# Observação x Estado

- **Observação:** informações do ambiente vêm de percepções do próprio agente
  - Robô capta informações por uma câmera
- **Estado:** informações vêm do próprio ambiente (maior precisão e quantidade de informações)
  - Robô recebe sua posição exata



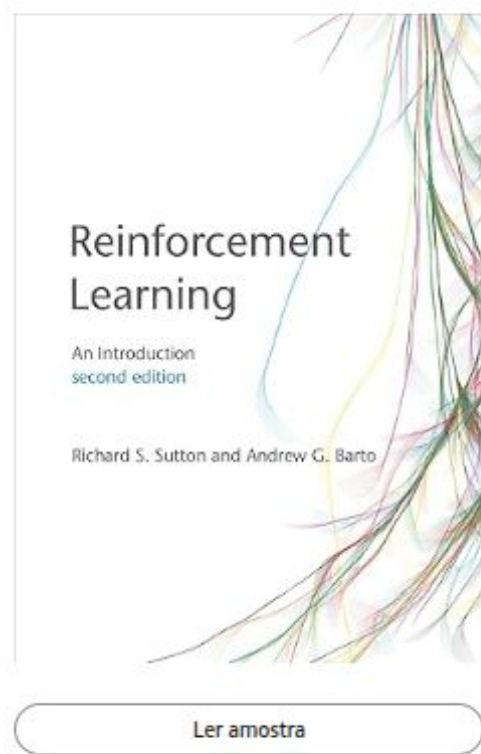
# Tipos de Ambientes

- **Totalmente Observável:** agente conhece tudo do ambiente (estado)
  - Caso ideal
  - Difícil de ocorrer
- **Parcialmente Observável:** agente conhece o ambiente de forma indireta (observação)
  - Caso mais comum



# Exemplo - Recomendação de livro

Livros > Livros Internacionais > Computação, Informática e Mídias Digitais > Programação > Algoritmos



## Reinforcement Learning, Second Edition: An Introduction Capa dura –

Ilustrado, 13 novembro 2018

Edição Inglês | por Richard S Sutton (Autor), Andrew G Barto (Autor)

4,6  583 avaliações de clientes

[Ver todos os formatos e edições](#)

**Economize 10%** nas compras de 2 ou mais [Comprar itens elegíveis >](#)

**Em até 10x R\$ 63,25 sem juros** [Ver parcelas disponíveis >](#)

**The significantly expanded and updated new edition of a widely used text on reinforcement learning, one of the most active research areas in artificial intelligence.**

Reinforcement learning, one of the most active research areas in artificial intelligence, is a computational approach to learning whereby an agent tries to maximize the total amount of reward it receives while interacting with a complex, uncertain environment. In *Reinforcement Learning*, Richard Sutton and Andrew Barto provide a clear and simple account of the field's key ideas and algorithms. This second edition has been significantly expanded and updated, presenting new topics and updating coverage of other topics.

Like the first edition, this second edition focuses on core online learning algorithms, with the more mathematical material set off in shaded boxes. Part I covers as much of reinforcement learning as possible without going beyond the tabular case for which exact solutions can be found. Many algorithms presented in this part are new to the second edition, including UCB, Expected Sarsa, and Double Learning. Part II extends these ideas to function approximation, with new sections on such topics as artificial neural networks and the Fourier basis, and offers expanded treatment of off-policy learning and policy-

[Leia mais](#)

 [Relatar um problema com este produto](#)

### Seguir o autor



Richard S.  
Sutton

[Seguir](#)

Número de  
páginas



552 páginas

Idioma



Inglês

Editora



Bradford Book

Data da publicação



13 novembro  
2018

Idade de leitura



18 anos e acima

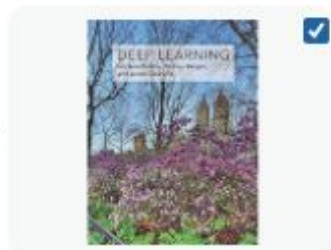


[Ver todos os detalhes](#)

### Frequentemente comprados juntos



+



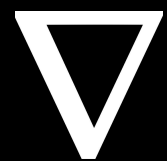
**Este item:** Reinforcement Learning, Second Edition: An Introduction  
R\$ 632<sup>32</sup>

Deep Learning  
R\$ 476<sup>00</sup>

Preço total: R\$ 1.108,92

[Adicionar ambos ao carrinho](#)





# Exemplo - Recomendação de livro

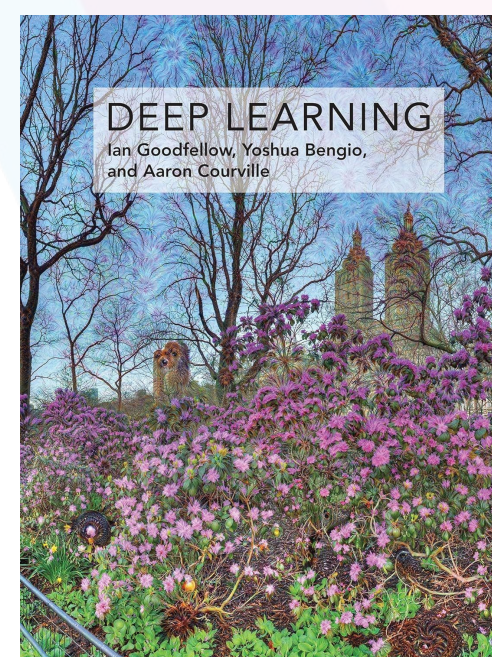
Livro comprado

## Reinforcement Learning

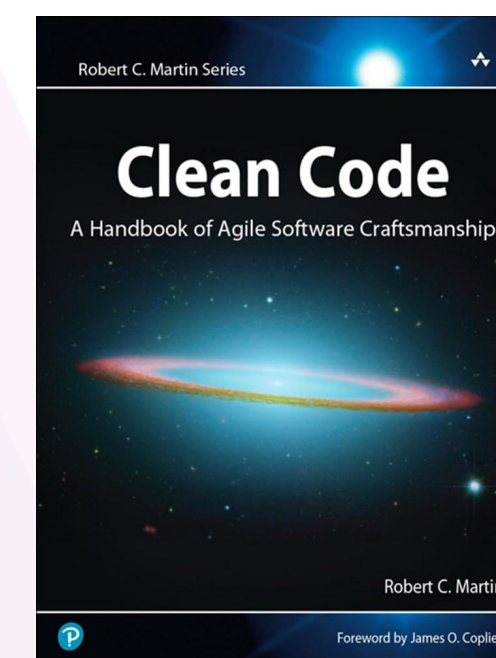
An Introduction  
second edition

Richard S. Sutton and Andrew G. Barto

opção 1

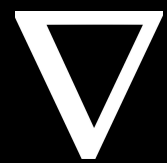


opção 2



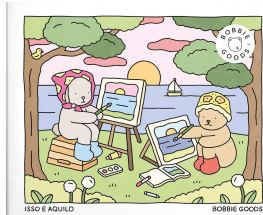
opção 3



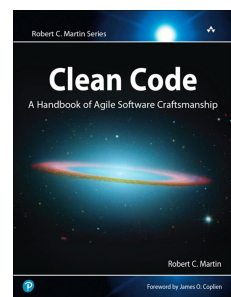


# Exemplo - Recomendação de livro

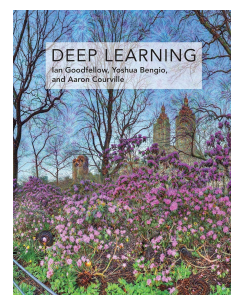
Livro  
sugerido  
opção 3



opção 2



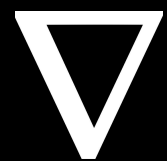
opção 1



vamos considerar a compra conjunta  
como 1, e a não compra como 0

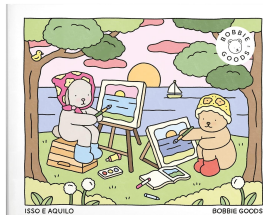
tentativas ao longo do tempo



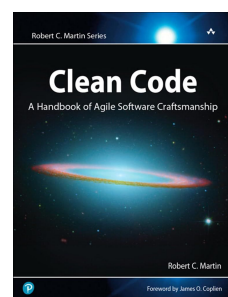


# Exemplo - Recomendação de livro

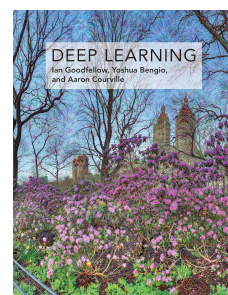
Livro  
sugerido  
opção 3



opção 2

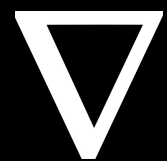


opção 1



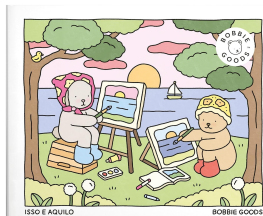
0

tentativas ao longo do tempo



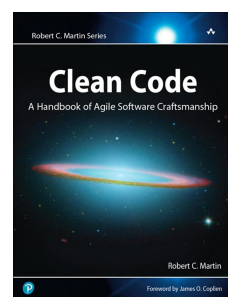
# Exemplo - Recomendação de livro

Livro  
sugerido  
opção 3



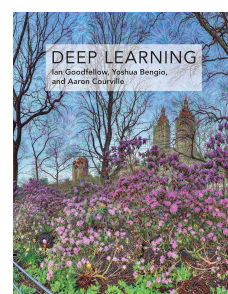
0

opção 2



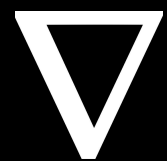
0

opção 1



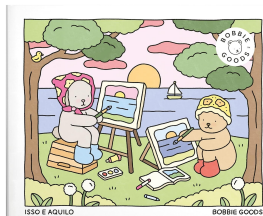
tentativas ao longo do tempo





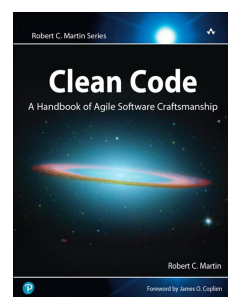
# Exemplo - Recomendação de livro

Livro  
sugerido  
opção 3



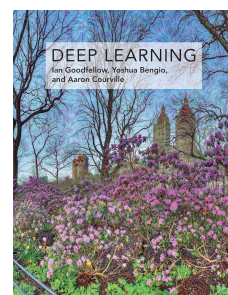
0

opção 2



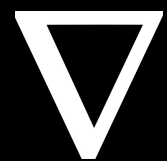
0

opção 1



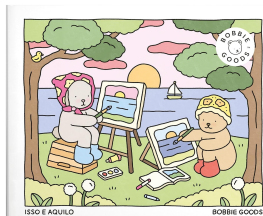
1

tentativas ao longo do tempo



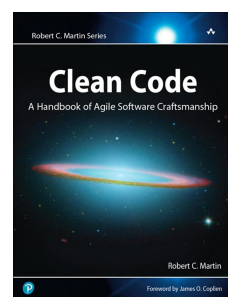
# Exemplo - Recomendação de livro

Livro  
sugerido  
opção 3



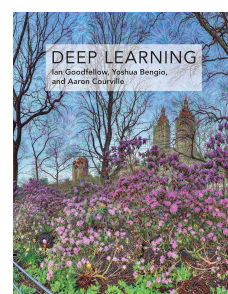
0

opção 2



0

opção 1



1

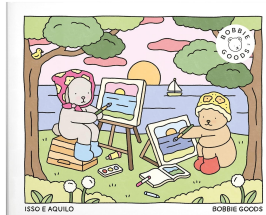
1

tentativas ao longo do tempo

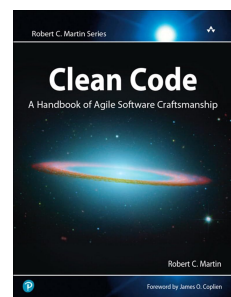


# Exemplo - Recomendação de livro

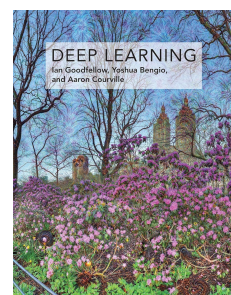
Livro  
sugerido  
opção 3



opção 2



opção 1



0

0

1

1

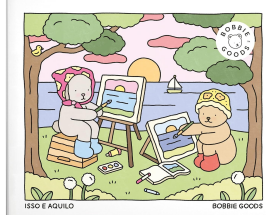
1

tentativas ao longo do tempo

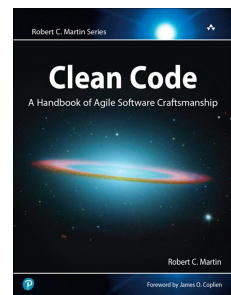


# Exemplo - Recomendação de livro

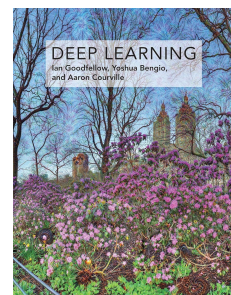
Livro  
sugerido  
opção 3



opção 2



opção 1



0

0

0

1

1

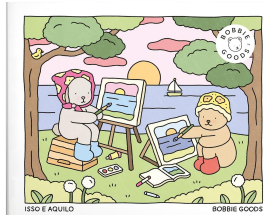
1

tentativas ao longo do tempo

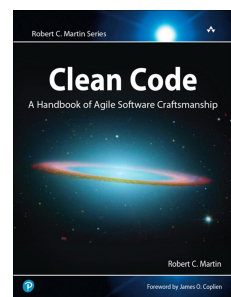


# Exemplo - Recomendação de livro

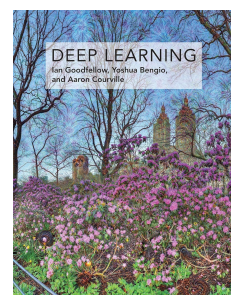
Livro  
sugerido  
opção 3



opção 2



opção 1



0

0

0

1

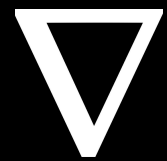
1

1

0

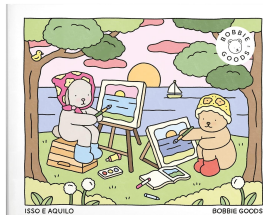
tentativas ao longo do tempo



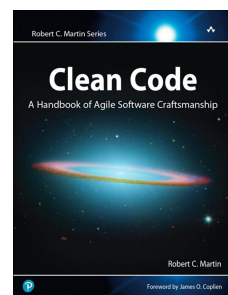


# Exemplo - Recomendação de livro

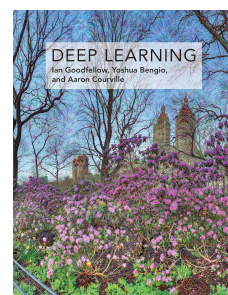
Livro  
sugerido  
opção 3



opção 2



opção 1



0

0

0

1

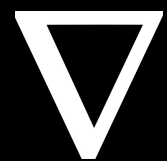
1

1

0

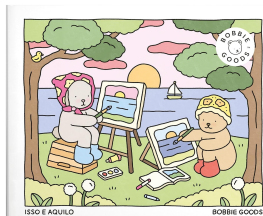
1

tentativas ao longo do tempo

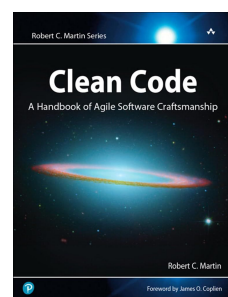


# Exemplo - Recomendação de livro

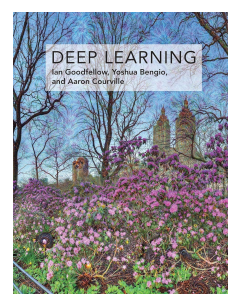
Livro  
sugerido  
opção 3



opção 2



opção 1



0

0

0

1

1

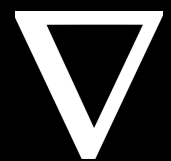
1

0

1

1

tentativas ao longo do tempo



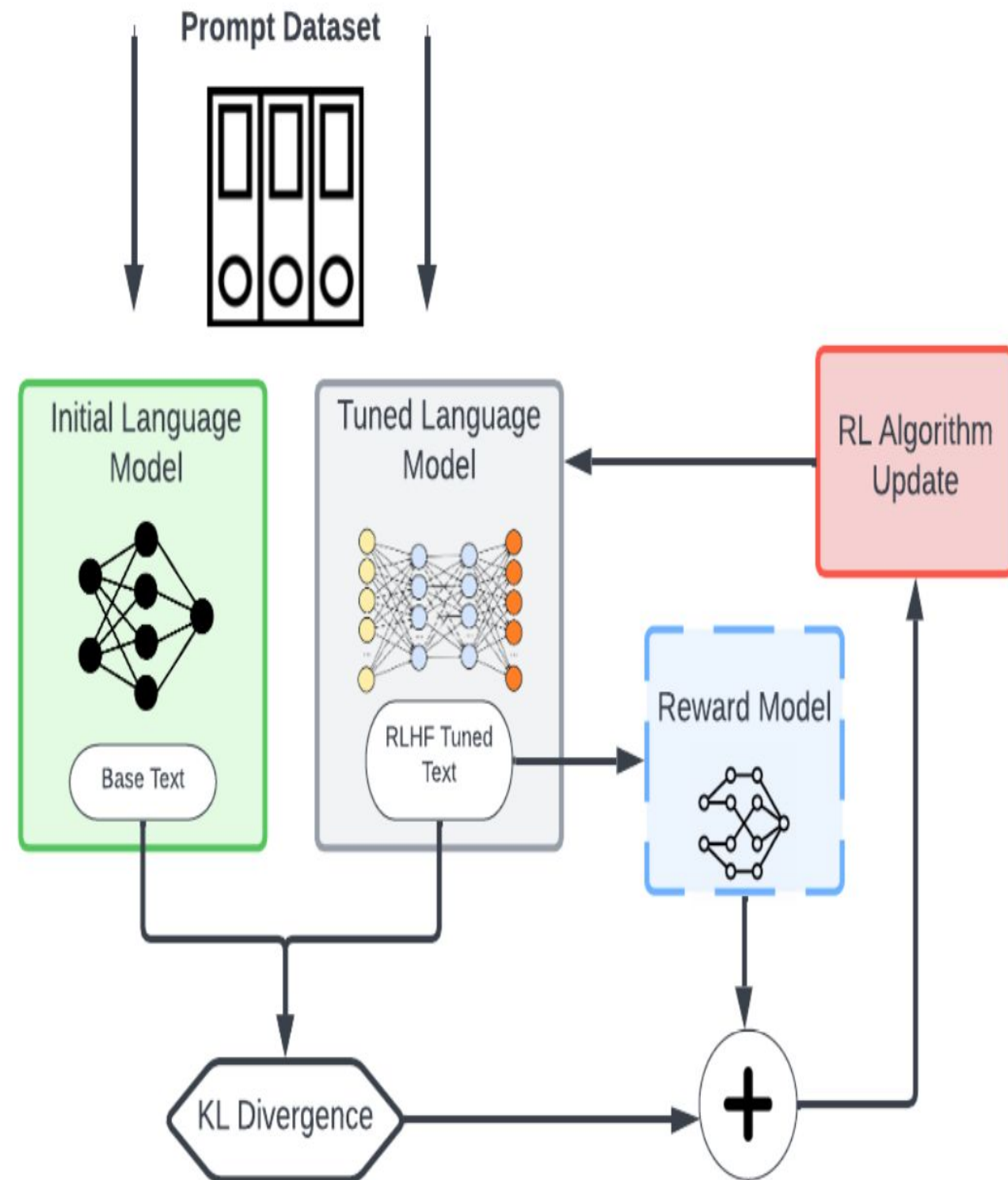
# Diferenças entre os paradigmas

Aspecto	Supervisionado	Não Supervisionado	Por Reforço (RL)
Precisa de rótulos?	Sim	Não	Não
Aprende com interação?	Não	Não	Sim
Utiliza recompensas?	Não	Não	Sim (inclusive atrasadas)
Objetivo principal	Prever saída correta	Descobrir estrutura	Maximizar recompensa total
Lida com decisões sequenciais?	Parcialmente	Não	Sim





# Aplicações - LLMs



**Modelo de recompensa:** Representa a preferência de usuários

**LLM** é treinado baseado no feedback humano das respostas geradas

**ChatGPT:** Modelo foi treinado utilizando essa técnica



# Aplicações - AWS DeepRacer

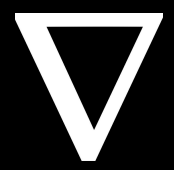


**Competição promovida pela AWS:** Carrinhos autônomos precisam percorrer uma pista de corrida o mais rápido possível

**Navegação autônoma:** Direção e Velocidade são controlados pelo modelo de RL

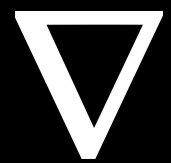
**Visão Computacional:** O estado do modelo em determinado momento é capturado por uma câmera





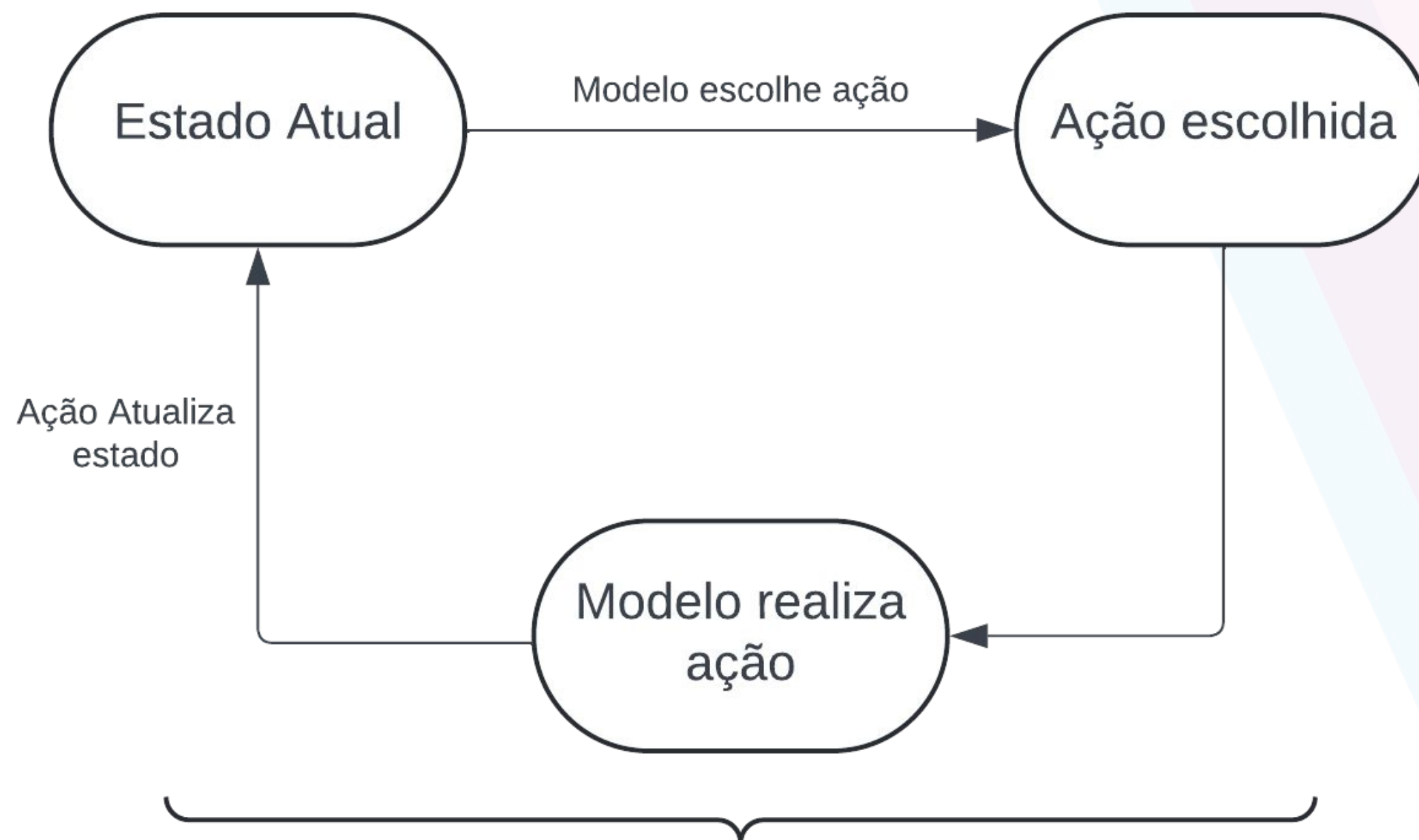
# Aplicações - AWS DeepRacer



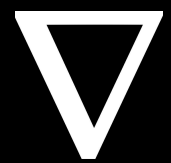


# Aplicações - AWS DeepRacer

Loop Execução

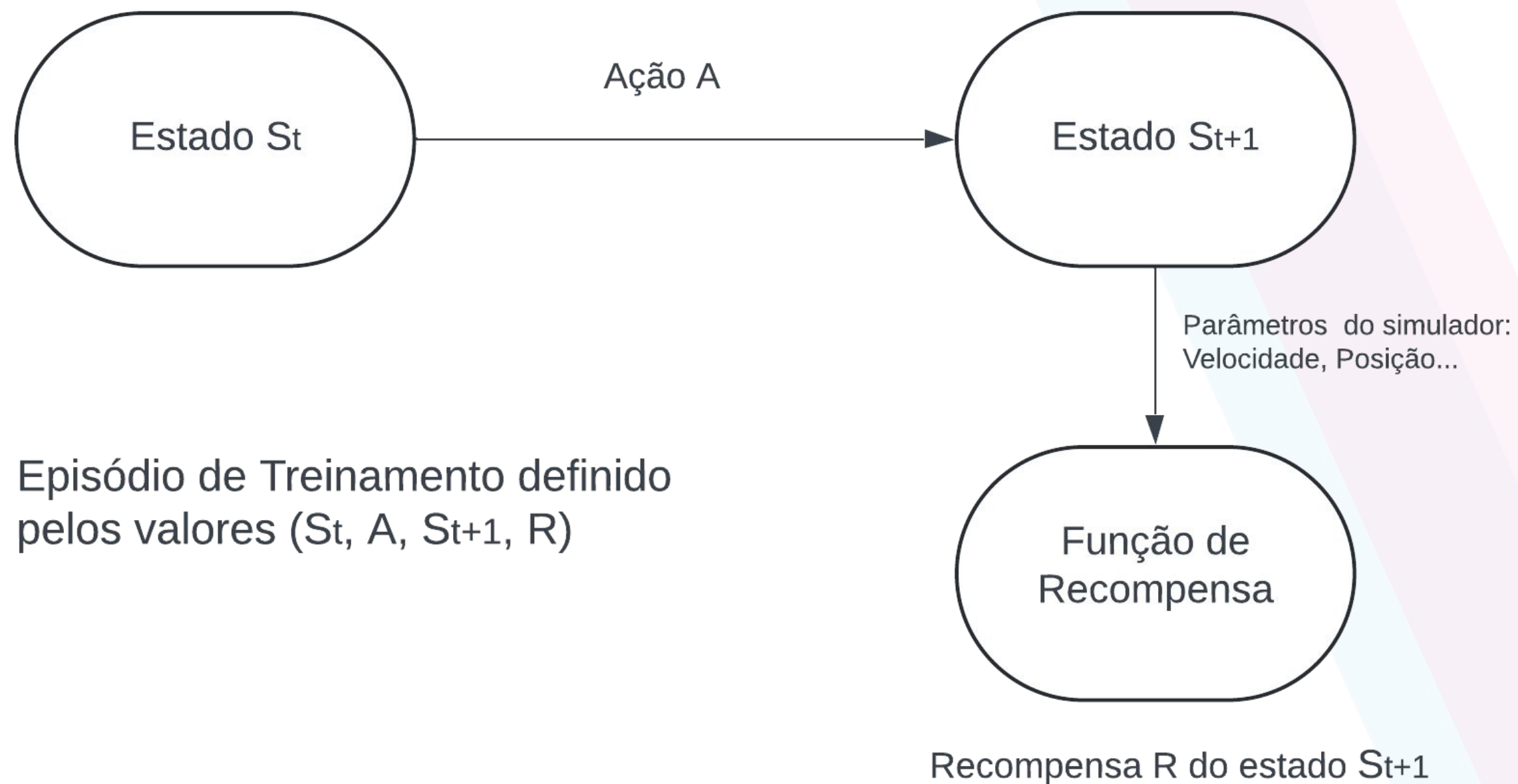


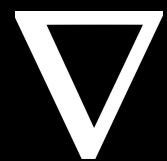
Dados treinamento: Estado atual, recompensa, ação e próximo estado



# Aplicações - AWS DeepRacer

## Treinamento





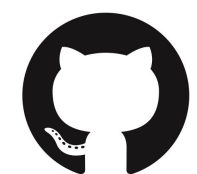
data@icmc.usp.br



@data.icmc



/c/DataICMC



/icmc-data



data.icmc.usp.br

|| obrigado!