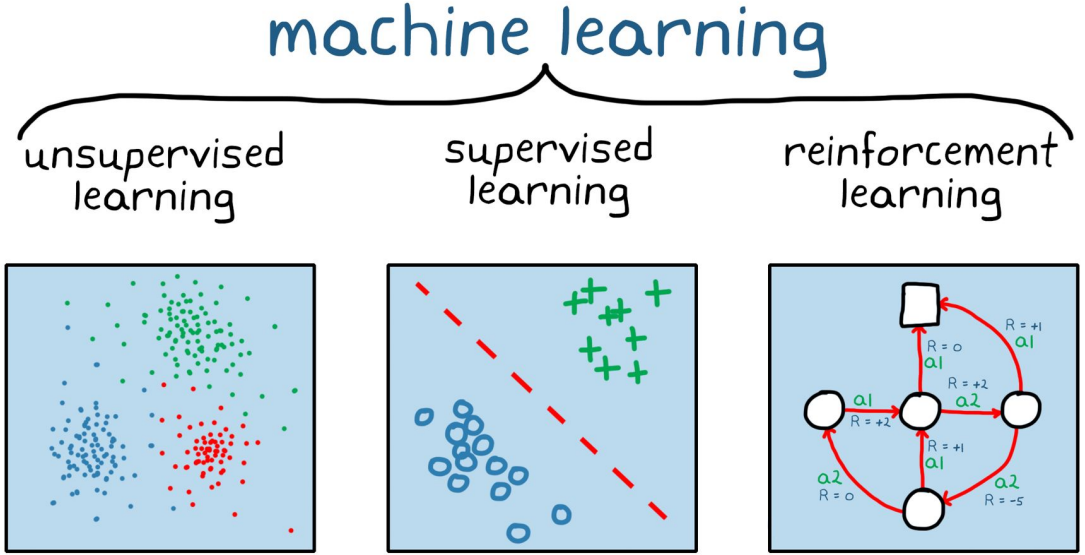
RL, 2025

Классификация областей ML

Обучение с подкреплением - набор методов решение задач принятия решений методом проб и ошибок



Основные понятия

Агент (Agent) — это система, которая принимает решения и выполняет действия в окружающей среде.

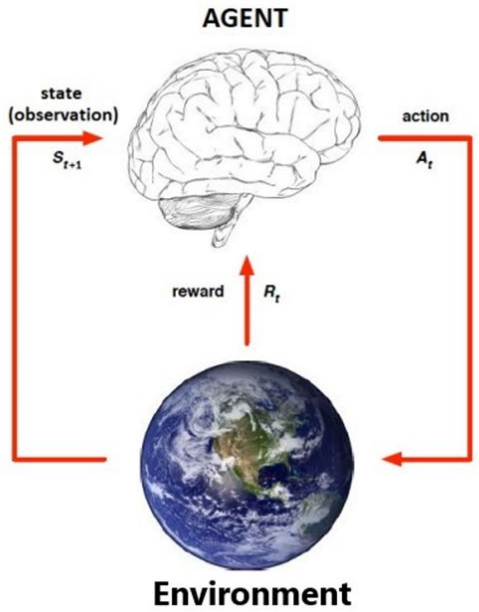
Среда (Environment) — это внешний мир, с которым агент взаимодействует.

Состояние (State) — это описание текущего положения агента в среде.

Действие (Action) — это выбор агента в каждом шаге взаимодействия с окружающей средой.

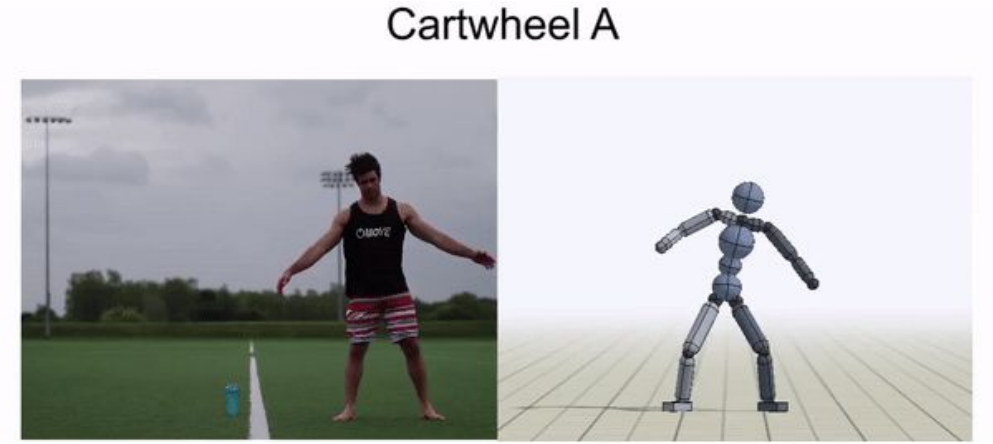
Политика (Policy) — это стратегия или правило, по которому агент выбирает действия в зависимости от состояния.

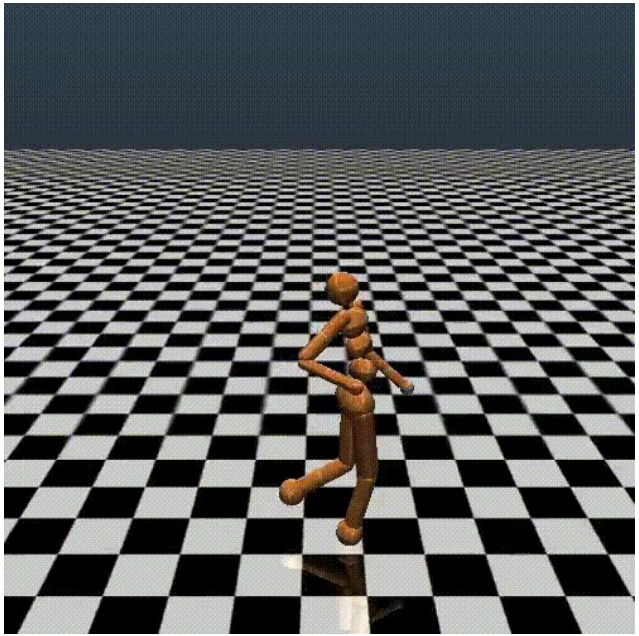
Награда (Reward) — это числовое значение, которое агент получает после выполнения действия в определенном состоянии. Награда отражает, насколько успешным было действие агента в конкретной ситуации. Агент стремится максимизировать суммарную награду.



Robotics (Locomtion)

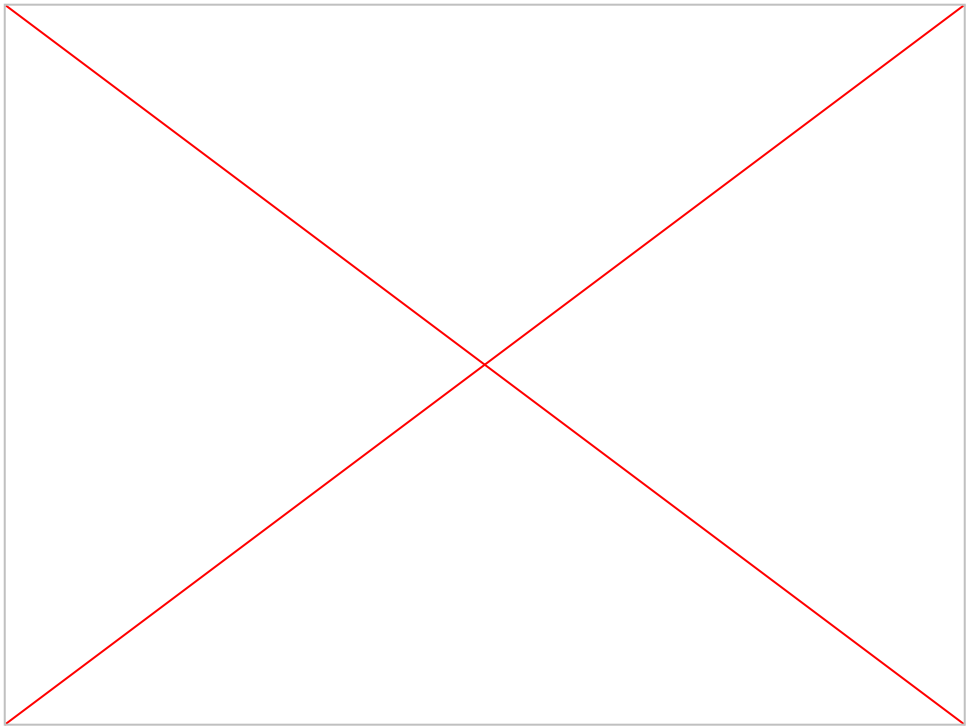
RL агент обучается в симуляции, а затем переносится на железо.





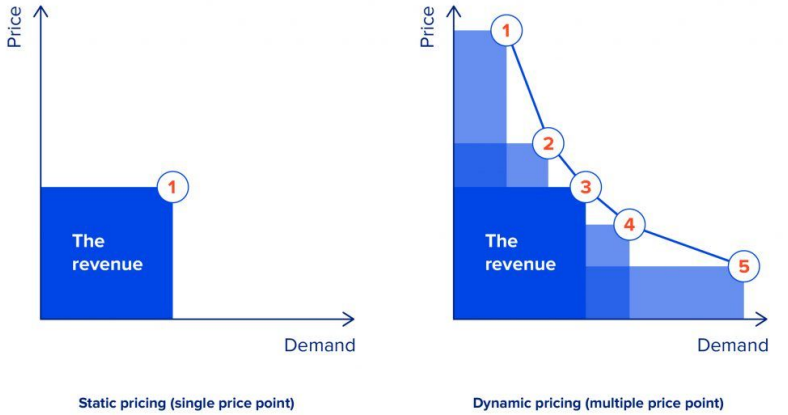
Robotics (Manipulation)

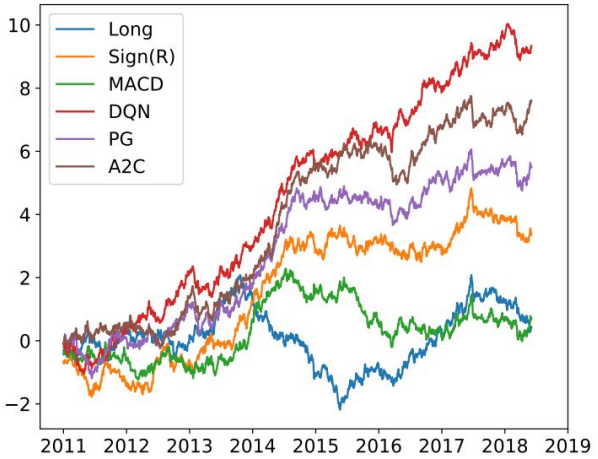
По экспертным демонстрациям работы манипулятора научиться повторять поведение (behavioral cloning)



Табличный ML

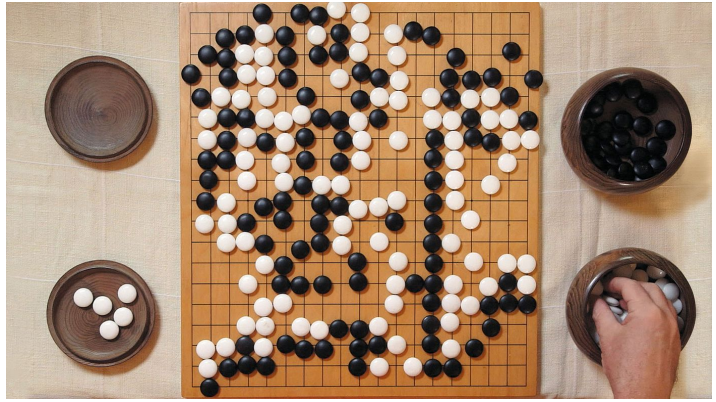
RL применяется в динамическом ценообразовании. Реже в трейдинге и рекомендательных системах.

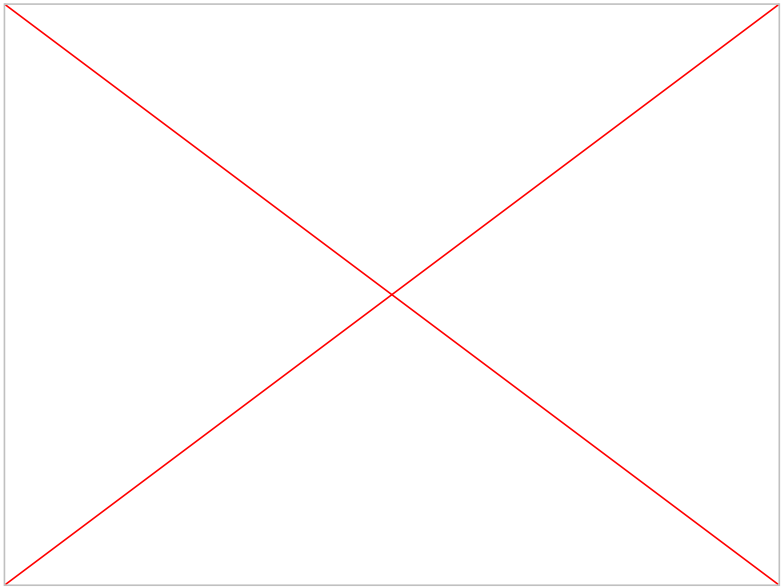


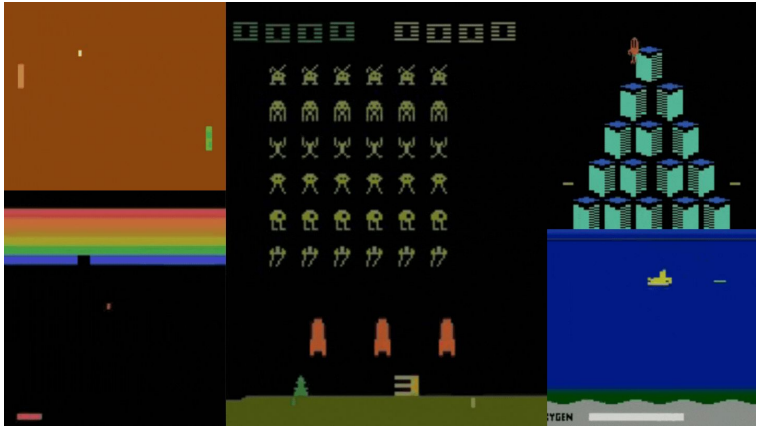


Игры и когнитивное поведение

RL прошел игры Atari, обыграл человека в Go, добыл алмаз в майнкрафте

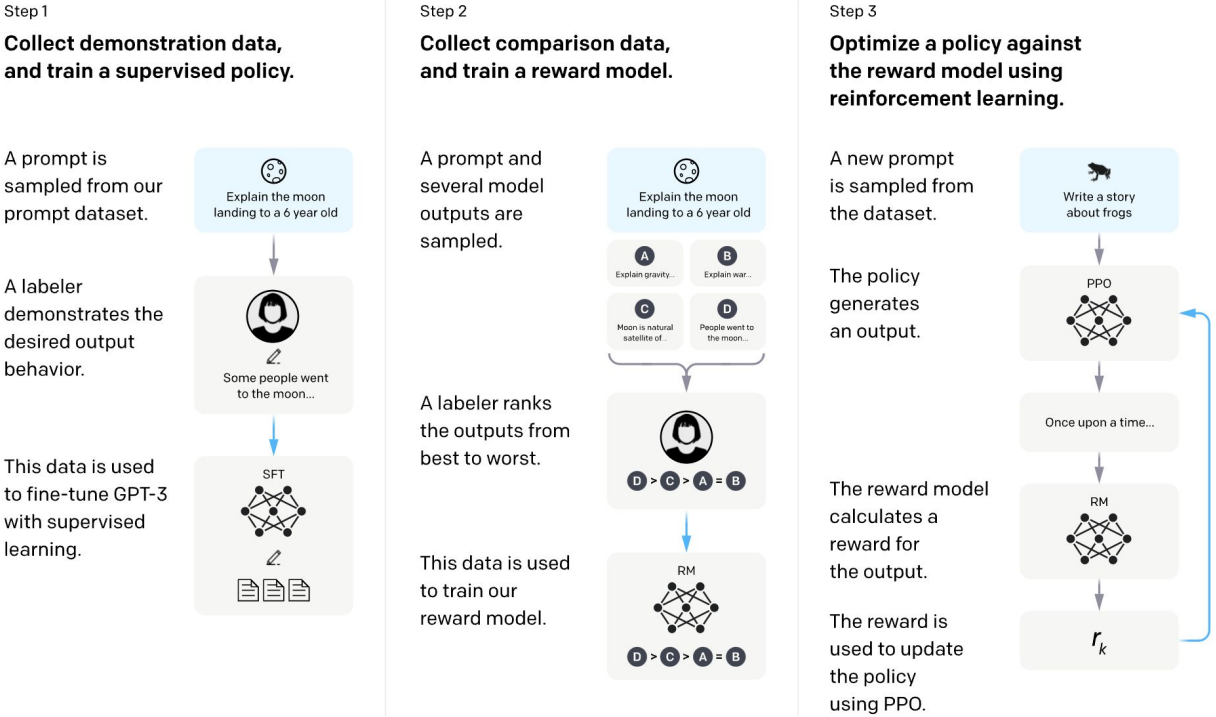






LLM (Alignement)

RL используется для улучшения качества ответов генеративных моделей.



LLM (Reasoning)

RL используется для улучшения когнитивных способностей LLM (по крайней мере так кажется)