**АЛГОРИТМЫ POLICY GRADIENT**

Подготовка разбора методов обучения с подкреплением семейства градиента политики (Policy Gradient), таких как ванильный PG, PPO, TRPO и других (количество алгоритмов выбирает докладчик, но чем обширнее исследование, тем лучше) и реализация их на практике с последующим сравнением их эффективности (в качестве полигона для испытаний предлагается использовать игры Atari или движок OpenAI от Gym)

Обучение с подкреплением – это тип алгоритма машинного обучения, который позволяет агенту учиться у окружающей среды методом проб и ошибок и использовать обратную связь от своих действий, чтобы узнать лучшую стратегию для достижения своих целей. Он работает путем назначения вознаграждений за успешные действия и наказаний за неуспешные, что позволяет агенту соответствующим образом корректировать свою стратегию.

**Градиентный алгоритм политики PG   
и лежащие в его основе принципы:**

Алгоритм Policy Gradient (PG) - это тип алгоритма обучения с подкреплением, используемый для повышения эффективности работы агента в конкретной задаче. Он работает на основе политики, определяющей, как агент должен действовать в той или иной ситуации. Затем агент получает обратную связь о своих действиях и использует ее для корректировки своей политики, чтобы в будущих ситуациях действовать лучше.

В основе алгоритма PG лежит идея градиента политики, который представляет собой скорость, с которой политика корректируется в ответ на наблюдаемое вознаграждение. Градиент политики рассчитывается путем взятия производной от ожидаемой прибыли в зависимости от параметров политики. Этот градиент затем используется для обновления параметров политики таким образом, чтобы агент мог максимизировать свою ожидаемую прибыль.

Затем агент использует эту обновленную политику для выполнения действий и наблюдения за вознаграждением, и процесс повторяется до тех пор, пока агент не выучит оптимальную политику. Этот процесс обучения называется итерацией политики. Ключевая идея PG заключается в том, что агента побуждают больше исследовать и рисковать, чтобы максимизировать ожидаемую прибыль, поскольку он может учиться на получаемых вознаграждениях.

Алгоритмы PG широко используются в робототехнике, видеоиграх и других приложениях ИИ. Они также используются в таких областях исследований, как обработка естественного языка и обучение с подкреплением.