

# Оглавление



<b>Предисловие</b> .....	<b>10</b>
<b>Вступление</b> .....	<b>12</b>
<b>Благодарности</b> .....	<b>14</b>
<b>О книге</b> .....	<b>17</b>
Для кого эта книга .....	17
Структура издания .....	17
О коде .....	18
<b>От издательства</b> .....	<b>19</b>
<b>Об авторе</b> .....	<b>20</b>
<b>Глава 1. Введение в глубокое обучение с подкреплением</b> .....	<b>21</b>
Что такое глубокое обучение с подкреплением .....	22
Прошлое, настоящее и будущее глубокого обучения с подкреплением .....	35
Целесообразность глубокого обучения с подкреплением .....	43
Определение четких обоюдных ожиданий .....	46
Подведем итоги .....	50
<b>Глава 2. Математические основы обучения с подкреплением</b> .....	<b>52</b>
Элементы обучения с подкреплением .....	54
MDP: двигатель среды .....	65
Подведем итоги .....	84

<b>Глава 3. Баланс краткосрочных и долгосрочных целей</b> .....	<b>86</b>
Цель агента, принимающего решения .....	87
Планирование оптимальных последовательностей действий .....	99
Подведем итоги .....	117
<b>Глава 4. Баланс между сбором и использованием информации</b> .....	<b>119</b>
Проблема интерпретации оценочной обратной связи .....	121
Стратегическое исследование .....	139
Подведем итоги .....	151
<b>Глава 5. Оценка поведения агента</b> .....	<b>153</b>
Учимся прогнозировать ценность политик .....	155
Прогноз на основе нескольких шагов .....	173
Подведем итоги .....	187
<b>Глава 6. Улучшение поведения агентов</b> .....	<b>189</b>
Анатомия агентов обучения с подкреплением .....	190
Оптимизация политик поведения .....	198
Разделение поведения и обучения .....	208
Подведем итоги .....	222
<b>Глава 7. Более действенные и эффективные способы достижения целей</b> .....	<b>224</b>
Улучшение политик с помощью достоверных целей .....	226
Агенты, которые взаимодействуют, обучаются и планируют .....	237
Подведем итоги .....	257
<b>Глава 8. Введение в ценностно ориентированное глубокое обучение с подкреплением</b> .....	<b>259</b>
Тип обратной связи, который используют агенты глубокого обучения с подкреплением .....	261
Введение в аппроксимацию функций для обучения с подкреплением ....	268
NFQ: первая попытка реализовать ценностно ориентированное глубокое обучение с подкреплением .....	273
Подведем итоги .....	294

<b>Глава 9. Более стабильные ценностно ориентированные методы</b> .....	<b>296</b>
DQN: делаем RL похожим на контролируемое обучение .....	297
Двойная DDQN: борьба с завышением прогнозов функций ценности действий .....	314
Подведем итоги .....	329
<b>Глава 10. Ценностно ориентированные методы с эффективным использованием выборок</b> .....	<b>332</b>
Дуэльная DDQN: архитектура нейросети, рассчитанная на обучение с подкреплением .....	333
PER: приоритетное воспроизведение полезного опыта .....	346
Подведем итоги .....	360
<b>Глава 11. Методы градиента политик и «актер-критик»</b> .....	<b>362</b>
REINFORCE: обучение политик на основе результатов .....	363
VPG: формирование функции ценности .....	374
A3C: параллельное обновление политики .....	380
GAE: надежное прогнозирование преимущества .....	387
A2C: синхронное обновление политик .....	391
Подведем итоги .....	399
<b>Глава 12. Продвинутые методы «актер-критик»</b> .....	<b>401</b>
DDPG: аппроксимация детерминированной политики .....	403
TD3: лучшие оптимизации для DDPG .....	410
SAC: максимизация ожидаемой выгоды и энтропии .....	417
PPO: ограничение этапа оптимизации .....	425
Подведем итоги .....	435
<b>Глава 13. Путь к сильному искусственному интеллекту</b> .....	<b>437</b>
Важные темы, которые были рассмотрены, и те, которые остались без внимания .....	439
Углубленные аспекты AGI .....	451
Что дальше? .....	458
Подведем итоги .....	461