

---

**УТВЕРЖДЕНА**

Решением Ученого совета  
АНО ВО «Центральный университет»  
«25» мая 2026 г.  
Протокол № 2

**Рабочая программа дисциплины (модуля)  
«Обучение с подкреплением»**

**Направление подготовки:** 02.04.01 Математика и компьютерные науки

**Направленность (профиль) подготовки:** Машинное обучение в индустрии

**Квалификация (степень) выпускника:** магистр

**Форма обучения:** очная

**Срок освоения программы:** 2 года

**Год набора:** 2026

**Москва  
2026**

## Содержание

|                                                     |    |
|-----------------------------------------------------|----|
| 1. Краткая характеристика дисциплины (модуля) ..... | 3  |
| 2. Перечень планируемых результатов обучения.....   | 5  |
| 3. Тематический план.....                           | 8  |
| 4. Содержание дисциплины (модуля).....              | 8  |
| 5. Учебно-методическое обеспечение .....            | 10 |
| 6. Материально-техническое обеспечение .....        | 10 |
| 7. Методические и оценочные материалы .....         | 12 |

## **1. Краткая характеристика дисциплины (модуля)**

Рабочая программа дисциплины (модуля) «Обучение с подкреплением» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – магистратура по направлению 02.04.01 Математика и компьютерные науки, профиль Машинное обучение в промышленности, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 810 от 23.08.2017 года.

Изучение дисциплины (модуля) «Обучение с подкреплением» позволяет сформировать у студентов глубокое понимание математических основ, алгоритмических подходов и практических методов обучения агентов в динамических средах, а также развить навыки формализации задач принятия решений, реализации и анализа алгоритмов обучения с подкреплением в условиях неопределённости, частичной наблюдаемости и сложной структуры вознаграждений.

### **Место дисциплины (модуля) в структуре образовательной программы**

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки магистратуры по направлению 02.04.01 Математика и компьютерные науки, профиль Машинное обучение в промышленности и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений как дисциплина по выбору.

Дисциплина (модуль) доступна к изучению на 2 курсе в 3 семестре. Доступна для прохождения при условии успешного завершения дисциплины (модуля) «Глубокое обучение (Deep Learning)».

**Цель изучения дисциплины (модуля):** подготовить обучающихся к решению задач последовательного принятия решений с использованием методов обучения с подкреплением, обеспечив их теоретической базой и практическими навыками для разработки, реализации, оценки и анализа агентов, способных эффективно взаимодействовать со средой и адаптироваться к изменяющимся условиям.

### **Задачи изучения дисциплины (модуля):**

- освоить математическую формализацию задач обучения с подкреплением: MDP, POMDP, функции ценности, уравнения Беллмана, стратегии дисконтирования;
- изучить ключевые алгоритмы: от классических (динамическое программирование, Monte Carlo, TD) до современных (DQN, PPO, DDPG, SAC);
- научиться формализовать прикладные задачи (робототехника, игры, управление) как процессы принятия решений и выбирать адекватный метод обучения;
- приобрести навыки реализации агентов на Python с использованием современных библиотек (Gymnasium, PyTorch) и оценки их производительности по ключевым метрикам;
- развить компетенции в проведении воспроизводимых экспериментов, анализе траекторий, политик и функций ценности, а также подборе и валидации гиперпараметров.

### **В результате освоения дисциплины (модуля) обучающийся должен:**

#### **знать:**

- математическую постановку задач обучения с подкреплением: марковские

процессы принятия решений и частично наблюдаемые марковские процессы принятия решений, функции ценности, политики поведения агента, уравнения Беллмана, постановки с дисконтированным и средним вознаграждением;

- основные алгоритмы обучения с подкреплением: динамическое программирование, методы Монте-Карло, обучение с временными разностями, обучение Q-функции, глубокие Q-сети, методы с оцениванием распределения суммарного вознаграждения, методы градиента политики, методы с отдельным обучением политики и функции ценности, проксимальная оптимизация политики, глубокий детерминированный градиент политики, мягкие методы с отдельным обучением политики и функции ценности, обучение с использованием модели среды, обучение по демонстрациям, поиск по дереву методом Монте-Карло, обучение с подкреплением по обратной связи от человека.

***уметь:***

- формализовать прикладную задачу как марковский или частично наблюдаемый марковский процесс принятия решений и выбирать подходящий алгоритм обучения;
- реализовывать агентов обучения с подкреплением на языке Python с использованием библиотек Gymnasium, PyTorch и NumPy, а также оценивать качество политик по суммарному вознаграждению, эффективности использования данных, стабильности обучения и устойчивости к изменению условий.

***владеть:***

- практиками проведения воспроизводимого эксперимента по обучению с подкреплением: фиксацией генераторов случайных чисел, ведением журнала экспериментов, использованием единого протокола оценки и подбором гиперпараметров;
- навыками анализа траекторий взаимодействия агента со средой, функций ценности и политик поведения агента.

## 2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

| Компетенция | Содержание компетенции                                                                                                      | Индикатор компетенции | Перечень планируемых результатов обучения по дисциплине (модулю)                                                                                                                                                                      |
|-------------|-----------------------------------------------------------------------------------------------------------------------------|-----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| УК-6.       | Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки | УК-6.1.               | Знает основные методы самооценки и анализа своей деятельности, а также принципы управления временем и целеполагания                                                                                                                   |
|             |                                                                                                                             | УК-6.2.               | Умеет ставить реалистичные и достижимые цели, определять приоритеты в своей деятельности, а также разрабатывать и внедрять планы по совершенствованию своих навыков и компетенций на основе полученной самооценки                     |
|             |                                                                                                                             | УК-6.3.               | Имеет практический опыт применения методов самооценки в своей профессиональной деятельности, включая участие в тренингах, семинарах и проектах, направленных на развитие личной эффективности и профессионального роста               |
| ОПК-1.      | Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и компьютерной математики               | ОПК-1.1.              | Знает основные методы и подходы к решению задач прикладной и компьютерной математики, включая алгоритмы, математическое моделирование и теорию оптимизации, а также современные инструменты и технологии, используемые в этой области |
|             |                                                                                                                             | ОПК-1.2.              | Умеет анализировать и формулировать математические задачи, применять соответствующие методы и алгоритмы для их решения, а также интерпретировать и                                                                                    |

|       |                                                                                                                         |          |                                                                                                                                                                                                                               |
|-------|-------------------------------------------------------------------------------------------------------------------------|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|       |                                                                                                                         |          | представлять результаты в понятной и доступной форме                                                                                                                                                                          |
|       |                                                                                                                         | ОПК-1.3. | Имеет практический опыт работы над проектами или исследованиями в области прикладной и компьютерной математики, включая участие в конкурсах, олимпиадах или научных публикациях, где были решены актуальные и значимые задачи |
| ПК-1. | Способен определять общие формы и закономерности области машинного обучения в индустрии                                 | ПК-1.1.  | Знает основные теоретические концепции и принципы, относящиеся к области машинного обучения в индустрии, а также ключевые закономерности и модели, которые помогают в анализе и интерпретации данных                          |
|       |                                                                                                                         | ПК-1.2.  | Умеет проводить систематический анализ области разработки, выявлять и формулировать общие закономерности и тенденции, а также применять методы исследования для получения новых знаний и понимания                            |
|       |                                                                                                                         | ПК-1.3.  | Имеет практический опыт работы в области машинного обучения в индустрии, включая участие в научных проектах, исследованиях или практических заданиях, где были выявлены и описаны общие формы и закономерности                |
| ПК-3. | Способен решать задачи профессиональной деятельности, формулировать результат, увидеть следствия полученного результата | ПК-3.1.  | Знает основные принципы и методы решения задач профессиональной деятельности, а также способы формулирования и представления результатов, включая анализ последствий и их значимость в контексте проекта                      |
|       |                                                                                                                         | ПК-3.2.  | Умеет применять математические и компьютерные методы для решения конкретных задач,                                                                                                                                            |

|       |                                                                                                                    |         |                                                                                                                                                                                                                                |
|-------|--------------------------------------------------------------------------------------------------------------------|---------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|       |                                                                                                                    |         | формулировать четкие и обоснованные результаты, а также анализировать их последствия для дальнейших действий и решений                                                                                                         |
|       |                                                                                                                    | ПК-3.3. | Имеет практический опыт в решении профессиональных задач, включая участие в проектах, где были получены результаты и проанализированы их следствия, что способствовало принятию обоснованных решений                           |
| ПК-6. | Способен разрабатывать программное обеспечение для решения прикладных задач в сфере машинного обучения в индустрии | ПК-6.1. | Знает основные языки программирования, методы разработки программного обеспечения, а также принципы проектирования и архитектуры программных систем, применяемых в машинном обучении в индустрии                               |
|       |                                                                                                                    | ПК-6.2. | Умеет анализировать прикладные задачи, разрабатывать алгоритмы и реализовывать их в виде программного обеспечения, используя современные инструменты и технологии, а также проводить тестирование и отладку созданных решений  |
|       |                                                                                                                    | ПК-6.3. | Имеет практический опыт разработки программного обеспечения в рамках реальных проектов, включая участие в командах, где были успешно реализованы решения для конкретных прикладных задач в сфере профессиональной деятельности |

### 3. Тематический план

| №<br>п/п                               | Наименование<br>раздела дисциплины<br>(модуля)           | Трудоемкость, академические часы |                                           |          |                               | ТКУ<br>(текущий<br>контроль<br>успеваемости) |
|----------------------------------------|----------------------------------------------------------|----------------------------------|-------------------------------------------|----------|-------------------------------|----------------------------------------------|
|                                        |                                                          | Очная форма                      |                                           |          |                               |                                              |
|                                        |                                                          | Аудиторная работа                |                                           | Контроль | Самостояте<br>льная<br>работа |                                              |
|                                        |                                                          | Лекции                           | Семинары<br>(практичес<br>кие<br>занятия) |          |                               |                                              |
| 1                                      | Математические основы RL                                 | 6                                | 6                                         |          | 26                            | Домашние задания                             |
| 2                                      | Value-based методы и исследование среды                  | 6                                | 6                                         |          | 25                            | Домашние задания                             |
| 3                                      | Policy gradient и actor-critic методы                    | 6                                | 6                                         |          | 25                            | Домашние задания                             |
| 4                                      | Планирование, model-based RL и обучение по демонстрациям | 6                                | 6                                         |          | 25                            | Домашние задания                             |
| 5                                      | Современные приложения и проектная работа                | 6                                | 6                                         |          | 25                            | Домашние задания<br>Мини-проект              |
|                                        | Экзамен                                                  |                                  |                                           | 4        |                               |                                              |
| Итого:                                 |                                                          | 30                               | 30                                        | 4        | 126                           |                                              |
| Объем дисциплины (модуля) (в ак. ч.)   |                                                          | 190                              |                                           |          |                               |                                              |
| Объем дисциплины (модуля) (в зач. ед.) |                                                          | 5                                |                                           |          |                               |                                              |

### 4. Содержание дисциплины (модуля)

| №п/п | Наименование раздела дисциплины<br>(модуля) | Содержание дисциплины (модуля) по темам                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------|---------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | Математические основы RL                    | Введение в обучение с подкреплением как в задачу последовательного принятия решений: кросс-энтропийный метод, среды для моделирования и проведения экспериментов<br>Марковские процессы принятия решений: уравнения Беллмана, динамическое программирование, итерация по ценности и итерация по стратегии<br>Табличные методы обучения с подкреплением: методы Монте-Карло, обучение с временными разностями, методы оценки функции действия-ценности, баланс исследования и использования |
| 2    | Value-based методы и исследование среды     | Глубокие методы оценки функции действия-ценности: буфер опыта, целевая сеть, устойчивость обучения и архитектурные модификации<br>Распределительное обучение с подкреплением: моделирование распределения отдачи и квантильные методы                                                                                                                                                                                                                                                      |

|   |                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                          |
|---|----------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |                                                          | Многорукие бандиты и стратегии исследования: методы верхней доверительной границы, сэмплирование Томпсона, внутренняя мотивация                                                                                                                                                                                                                                                                                          |
| 3 | Policy gradient и actor-critic методы                    | Методы градиента стратегии: оценка градиента, базовая функция, преимущество, актор-критик<br>Методы доверительного обновления стратегии и обобщенная оценка преимущества<br>Непрерывное управление: методы актор-критик для непрерывных пространств действий, детерминированные и энтропийно-регуляризованные подходы                                                                                                    |
| 4 | Планирование, model-based RL и обучение по демонстрациям | Имитационное обучение и обратное обучение с подкреплением: клонирование поведения, восстановление функции награды, состязательные методы имитационного обучения<br>Поиск по дереву Монте-Карло и нейросетевые игровые системы: поиск, сеть ценности, сеть стратегии<br>Модельно-ориентированное обучение с подкреплением и планирование: линейно-квадратичное управление, обучение моделей среды, внутренние модели мира |
| 5 | Современные приложения и проектная работа                | Обучение с подкреплением по обратной связи от человека: моделирование награды, обучение по предпочтениям, регуляризация политики<br>Частично наблюдаемые марковские процессы принятия решений: политики с памятью и рекуррентные агенты                                                                                                                                                                                  |

## 5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

### *Основная литература:*

1. Уиндер, Ф. Обучение с подкреплением для реальных задач : практическое пособие / Ф. Уиндер. - Санкт-Петербург : БХВ-Петербург, 2022. - 400 с. - ISBN 978-5-9775-6885-2. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2140226>.

2. Плас, Дж. В. Python для сложных задач: наука о данных и машинное обучение : практическое руководство / Дж. В. Плас. - Санкт-Петербург : Питер, 2021. - 576 с. - (Серия «Бестселлеры O'Reilly»). - ISBN 978-5-4461-0914-2. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/1739601>.

### *Дополнительная литература:*

1. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : учебник / П. Флах. - 2-е изд. - Москва.:ДМК Пресс, 2023. - 401 с. - ISBN 978-5-89818-300-4. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2085038>.

## 6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;

— специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

| №  | Наименование портала<br>(издания, курса, документа)                            | Ссылка                                                                          |
|----|--------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| 1. | Научная электронная библиотека elibrary.ru библиотека                          | <a href="https://elibrary.ru/defaultx.asp">https://elibrary.ru/defaultx.asp</a> |
| 2. | База данных для IT-специалистов                                                | <a href="https://habr.com">https://habr.com</a>                                 |
| 3. | База данных ScienceDirect                                                      | <a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>       |
| 4. | Официальный сайт Министерства науки и высшего образования Российской Федерации | <a href="https://minobrnauki.gov.ru/">https://minobrnauki.gov.ru/</a>           |
| 5. | Федеральный портал «Российское образование»                                    | <a href="https://www.edu.ru/">https://www.edu.ru/</a>                           |
| 6. | Информационная система "Единое окно доступа к образовательным ресурсам"        | <a href="http://window.edu.ru/">http://window.edu.ru/</a>                       |
| 7. | Единая коллекция цифровых образовательных ресурсов                             | <a href="http://school-collection.edu.ru/">http://school-collection.edu.ru/</a> |
| 8. | Федеральный центр информационно - образовательных ресурсов                     | <a href="http://fcior.edu.ru/">http://fcior.edu.ru/</a>                         |

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

| Наименование ПО                                                  | Производство  | Лицензионное / свободно распространяемое |
|------------------------------------------------------------------|---------------|------------------------------------------|
| <b>Операционные системы:</b>                                     |               |                                          |
| Microsoft Imagine (Windows Client, Server)                       | зарубежное    | лицензионное                             |
| <b>Браузеры:</b>                                                 |               |                                          |
| Яндекс.Браузер                                                   | отечественное | свободно распространяемое                |
| Google Chrome                                                    | зарубежное    | свободно распространяемое                |
| <b>Офисные приложения:</b>                                       |               |                                          |
| Microsoft Imagine (Visio, OneNote)                               | зарубежное    | лицензионное                             |
| TeXstudio                                                        | зарубежное    | свободно распространяемое                |
| Adobe Acrobat Reader                                             | зарубежное    | свободно распространяемое                |
| <b>Программное обеспечение для планирования и учета времени:</b> |               |                                          |
| Toggle app                                                       | зарубежное    | свободно распространяемое                |
| <b>Системы управления проектами:</b>                             |               |                                          |
| Microsoft Imagine (Project)                                      | зарубежное    | лицензионное                             |

|                                                                     |               |                           |
|---------------------------------------------------------------------|---------------|---------------------------|
| <b>Системы управления базами данных:</b>                            |               |                           |
| Microsoft Imagine (SQL Server)                                      | зарубежное    | лицензионное              |
| <b>Системы резервного копирования (backup):</b>                     |               |                           |
| Acronis Backup Advanced for HyperV                                  | зарубежное    | лицензионное              |
| <b>Справочно-правовые системы:</b>                                  |               |                           |
| КонсультантПлюс: справочно-правовая система                         | отечественное | лицензионное              |
| <b>Средства антивирусной защиты:</b>                                |               |                           |
| Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition | отечественное | лицензионное              |
| <b>Среды разработки:</b>                                            |               |                           |
| Visual Studio Code                                                  | зарубежное    | свободно распространяемое |
| Bash (Unix shell)                                                   | зарубежное    | свободно распространяемое |
| Anaconda                                                            | зарубежное    | свободно распространяемое |
| Robotic Operating System                                            | зарубежное    | свободно распространяемое |
| CopelliaSim                                                         | зарубежное    | свободно распространяемое |
| Google Colaboratory                                                 | зарубежное    | свободно распространяемое |
| <b>Пакеты программных средств и библиотек:</b>                      |               |                           |
| AutoPsy                                                             | зарубежное    | свободно распространяемое |
| Interactive Disassembler (IDA)                                      | зарубежное    | свободно распространяемое |
| <b>Системы управления библиографической информацией:</b>            |               |                           |
| Zotero                                                              | зарубежное    | свободно распространяемое |
| <b>Сервисы и службы:</b>                                            |               |                           |
| Bind                                                                | зарубежное    | свободно распространяемое |
| Docker                                                              | зарубежное    | свободно распространяемое |

## 7. Методические и оценочные материалы

### Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Обучение с подкреплением» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, домашние задания, мини-проект, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

*Лекция* – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

*Семинар* — это форма учебной деятельности, проводимая в учебном заведении под

руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

*Домашнее задание* – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал, использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

*Мини-проект* — это краткосрочная практическая работа в малых группах на семинарах, направленная на применение теоретических знаний для решения конкретных задач.

В ходе мини-проекта студенты проводят анализ проблемы, разрабатывают техническое решение и представляют результаты в виде отчета, модели или презентации. Такой формат способствует развитию навыков самостоятельной работы, критического мышления и командного взаимодействия.

*Самостоятельная работа* – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

### **Система оценивания результатов обучения по дисциплине (модулю)**

#### **Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Обучение с подкреплением»**

Оценивание уровня учебных достижений обучающихся по дисциплине (модулю) осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

**Промежуточная аттестация** по дисциплине (модулю) осуществляется в форме **экзамена**, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, если иное не указано в силлабусе дисциплины:

| Десятибалльная оценка | Пятибалльная оценка | Общая характеристика результата обучения по дисциплине (модулю)                                                                                                                                                                                                                 |
|-----------------------|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 10                    | Отлично             | Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину (модуль). Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, |
| 9                     | Отлично             |                                                                                                                                                                                                                                                                                 |
| 8                     | Отлично             |                                                                                                                                                                                                                                                                                 |

| Десятибалльная оценка | Пятибалльная оценка | Общая характеристика результата обучения по дисциплине (модулю)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-----------------------|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                       |                     | обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.                                                                                                                                                                                                                                                                             |
| 7                     | Хорошо              | Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами. |
| 6                     | Хорошо              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 5                     | Удовлетворительно   | Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.                                                                                                                                                                                                                                                                                                                                                                                                    |
| 4                     | Удовлетворительно   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 3                     | Не сдан             | Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| 2                     | Не сдан             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 1                     | Не сдан             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |

Дисциплина (модуль) «Обучение с подкреплением» оценивается следующим образом:

| Активность       | Вес | Количество | Описание                                                                                                               |
|------------------|-----|------------|------------------------------------------------------------------------------------------------------------------------|
| Домашние задания | 50% | 5          | Набор задач по темам недели                                                                                            |
| Мини-проект      | 10% | 1          | Краткосрочная практическая работа в малых группах на занятиях, направленная на решение конкретных задач                |
| Экзамен          | 40% | 1          | Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю) |

**Формула расчёта итоговой оценки по дисциплине (модулю) «Обучение с подкреплением»:**  $0,5 \times \text{среднее за домашние задания} + 0,1 \times \text{мини-проект} + 0,4 \times$

экзамен».

## Текущий контроль успеваемости обучающихся по дисциплине (модулю)

### Примерные домашние задания

#### Домашнее задание 1. Табличные методы RL: Monte Carlo и Temporal Difference

##### Задание:

1. Реализуйте среду GridWorld (5×5) с заданными препятствиями и терминальным состоянием.
2. Реализуйте:
  - Метод Монте-Карло с первым посещением (first-visit MC) для оценки функции ценности состояния
  - SARSA и Q-обучение (TD(0)) для оценки
3. Проведите сравнение:
  - Сходимость функций ценности
  - Влияние коэффициента обучения и -жадной стратегии
  - Количество шагов до сходимости
4. Визуализируйте:
  - Политику агента на разных этапах обучения
  - Изменение по эпизодам
5. Объясните различия в поведении MC и TD методов.

##### Требования:

- Код на Python (NumPy, Gymnasium-like interface)
- Графики сходимости
- Аналитический отчёт (1–2 стр.)

#### Домашнее задание 2. Deep Q-Network и исследование среды

##### Задание:

1. Используйте среду CartPole-v1 из Gymnasium.
2. Реализуйте DQN с ключевыми компонентами:
  - Опыт-буфер (replay buffer)
  - Целевая сеть (target network)
  - -жадная стратегия с затуханием
3. Обучите агента:
  - Зафиксируйте кривую суммарного вознаграждения по эпизодам
  - Проведите анализ устойчивости обучения
4. Добавьте улучшения:
  - Double DQN или Dueling DQN
  - Сравните с базовой версией
5. Проанализируйте:
  - Влияние размера буфера и частоты обновления целевой сети
  - Эффективность использования данных

##### Требования:

- Код с комментариями (PyTorch или JAX)
- Графики обучения

- Сравнение подходов и выводы об устойчивости

### **Домашнее задание 3. Policy Gradient и Actor-Critic в непрерывной среде**

#### **Задание:**

1. Используйте среду Pendulum-v1 (непрерывное пространство действий).
2. Реализуйте:
  - Простой REINFORCE с базовой функцией (baseline)
  - Агента A2C (Advantage Actor-Critic)
3. Обучите обе модели:
  - Сравните кривые вознаграждения
  - Оцените дисперсию градиентов
4. Реализуйте и протестируйте:
  - Обобщённую оценку преимущества (GAE)
  - Сравните качество политики
5. Проанализируйте:
  - Стабильность обучения
  - Влияние энтропийной регуляризации на исследование

#### **Требования:**

- Реализация на PyTorch
- Визуализация траекторий и политик
- Обсуждение компромисса между смещением и дисперсией

### **Примерное описание задания и критерии оценивания к мини-проекту**

#### **Цель:**

применить изученные методы RL для решения задачи в условиях частичной наблюдаемости, сравнить эффективность различных подходов и провести анализ поведения агентов.

#### **Формат:**

групповая работа (2–3 человека), выполнение на семинарском занятии (3–4 академических часа)

#### **Этапы выполнения:**

1. **Формализация задачи (30 мин)**
  - Выбор или создание простой среды POMDP (например, "тёмная комната" с маркерами)
  - Определение состояний, действий, наблюдений, награды
2. **Реализация агентов (90 мин)**
  - Агент 1: DQN с фиксированной историей наблюдений (как вход)
  - Агент 2: A2C с LSTM-слоем для учёта памяти
  - Обучение в симуляторе (упрощённая реализация)
3. **Оценка и сравнение (30 мин)**
  - Сравнение по суммарному вознаграждению
  - Анализ траекторий и устойчивости
  - Обсуждение, какой агент лучше справляется с неопределённостью

### Критерии оценивания:

- **Корректность постановки задачи**

Задача формализована как процесс принятия решений с чётким определением состояний, действий, наблюдений, функции награды и динамики среды. Учтены особенности частичной наблюдаемости (POMDP), если применимо.

- **Адекватность выбранных методов**

Использованные алгоритмы обучения с подкреплением соответствуют постановке задачи (например, применение рекуррентных сетей для учёта истории при частичной наблюдаемости, корректное использование off-policy или on-policy подходов).

- **Работоспособность и корректность реализации**

Код агентов выполняется без ошибок, корректно взаимодействует со средой, реализует заявленные алгоритмы (DQN, A2C, LSTM и т.д.). Используются ключевые компоненты: буфер опыта, целевая сеть, механизмы обновления политики.

- **Качество анализа результатов**

Проведено сравнение агентов по ключевым метрикам: суммарное вознаграждение, стабильность обучения, скорость сходимости. Проанализированы траектории поведения, выявлены сильные и слабые стороны подходов.

- **Интерпретация и обоснование выводов**

Выводы о поведении агентов обоснованы с точки зрения теории RL (например, объяснение преимущества актор-критика с памятью в условиях неопределённости). Обсуждаются ограничения модели и возможные улучшения.

### Задания для промежуточной аттестации по дисциплине (модулю)

| № п/п | Задание                                                                                 | Ответ                                                                                     | Компетенция |
|-------|-----------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|-------------|
| 1.    | Как спланировать изучение методов обучения с подкреплением для профессионального роста? | План обучения с этапами, сроками и целями по темам (MDP, алгоритмы, среды)                | УК-6        |
| 2.    | Как определить приоритеты при выборе RL-алгоритма для проекта?                          | Оценка требований к среде, действиям, доступным данным и вычислительным ресурсам          | УК-6        |
| 3.    | Как оценить свой текущий уровень навыков в обучении с подкреплением?                    | Самоанализ компетенций, тестирование на стандартных средах, обратная связь от экспертов   | УК-6        |
| 4.    | Как организовать процесс воспроизводимости RL-экспериментов?                            | Фиксация random seed, документирование гиперпараметров, версионирование кода              | УК-6        |
| 5.    | Как математически записывается уравнение Беллмана для Q-функции?                        | $Q(s,a) = R(s,a) + \gamma \cdot \max_{a'} Q(s',a')$<br>— рекурсивное соотношение ценности | ОПК-1       |
| 6.    | Как математически определяется advantage-функция?                                       | $A(s,a) = Q(s,a) - V(s)$ — превышение над средним                                         | ОПК-1       |

|     |                                                                      |                                                                            |       |
|-----|----------------------------------------------------------------------|----------------------------------------------------------------------------|-------|
|     |                                                                      | вознаграждением                                                            |       |
| 7.  | На каком математическом принципе основан алгоритм MCTS?              | Поиск по дереву с симуляцией траекторий и статистической оценкой узлов     | ОПК-1 |
| 8.  | Как математически формулируется баланс исследования и использования? | Через UCSB или вероятностное сэмплирование из апостериорного распределения | ОПК-1 |
| 9.  | Что характеризует MDP как модель в индустриальных приложениях RL?    | Марковское свойство: будущее зависит только от текущего состояния          | ПК-1  |
| 10. | Какие закономерности отличают on-policy от off-policy методов?       | SARSA обучается на текущей политике, Q-learning — на оптимальной           | ПК-1  |
| 11. | Что представляет собой PPO в контексте современных методов RL?       | Проксимальная оптимизация политики с ограничением на шаг обновления        | ПК-1  |
| 12. | Какие тенденции наблюдаются в методах для непрерывных действий?      | Использование детерминированных политик (DDPG, SAC) с актором-критиком     | ПК-1  |
| 13. | Какую задачу решает replay buffer в архитектуре DQN?                 | Декорреляция данных и эффективное переиспользование опыта для стабильности | ПК-3  |
| 14. | Каковы последствия недостаточного исследования (exploration) в RL?   | Застревание в локальных оптимумах, неоптимальная политика                  | ПК-3  |
| 15. | Как решить проблему нестабильности обучения в глубоком RL?           | Использование целевой сети, нормализации, clipping градиентов              | ПК-3  |
| 16. | Какое решение позволяет целевая сеть в DQN-архитектуре?              | Стабилизация обучения за счёт замороженных весов для расчёта targets       | ПК-3  |
| 17. | Как программно реализовать Q-learning алгоритм?                      | Цикл обновления Q-таблицы или нейросети с формулой Беллмана                | ПК-6  |
| 18. | Какие библиотеки используются для разработки RL-агентов?             | Stable Baselines3, RLlib, Gym/Gymnasium, PyTorch, TensorFlow               | ПК-6  |
| 19. | Как тестировать корректность реализации RL-агента?                   | Проверка на простых средах, сходимость к известным решениям, юнит-тесты    | ПК-6  |
| 20. | Как оптимизировать код обучения RL-модели?                           | Векторизация, батчевая обработка, использование GPU, профилирование        | ПК-6  |