

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 22.05.2026 11:38:35
Уникальный программный ключ:
ca953a0120d891083f939673078ef1a989dae18a

**Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы»
Факультет искусственного интеллекта**

(наименование основного учебного подразделения (ОУП) – разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

02.04.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

УПРАВЛЕНИЕ ДАННЫМИ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

(наименование (профиль/специализация) ОП ВО)

2026 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Обучение с подкреплением» входит в программу магистратуры «Управление данными и искусственный интеллект» по направлению 02.04.02 «Фундаментальная информатика и информационные технологии» и изучается в 3 семестре 2 курса. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 2 разделов и 12 тем и направлена на изучение развитие у студентов как фундаментальных, так и современных компетенций в области проектирования, анализа, объяснения и внедрения RL-алгоритмов для сложных задач управления и оптимизации. Программа выходит за рамки базовых алгоритмов, делая акцент на Deep RL, устойчивости, автоматизации экспериментов, объяснимости решений и интеграции RL в мультиагентные и отраслевые системы. Целью освоения дисциплины является дать студентам углубленные знания по теории и практике RL, научить их анализировать сходимость и стабильность моделей, реализовывать и тестировать сложные RL-решения в реальных средах, обеспечивать интерпретируемость, учитывать вопросы безопасности и этики.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Обучение с подкреплением» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
УК-2	Способен управлять проектом на всех этапах его жизненного цикла	УК-2.1 Знает необходимые для осуществления профессиональной деятельности правовые нормы; УК-2.2 Умеет определять круг задач в рамках избранных видов профессиональной деятельности, планировать собственную деятельность исходя из имеющихся ресурсов; соотносить главное и второстепенное, решать поставленные задачи в рамках избранных видов профессиональной деятельности;
ОПК-1	Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	ОПК-1.1 Обладает фундаментальными знаниями в области математических и естественных наук, теории коммуникаций; ОПК-1.2 Умеет осуществлять первичный сбор и анализ материала, интерпретировать различные математические объекты;
ОПК-5	Способен устанавливать и сопровождать программное обеспечение информационных систем, осуществлять эффективное управление разработкой программных средств и проектов	ОПК-5.2 разрабатывает, внедряет и сопровождает программное обеспечение информационных систем;
ПК-2	Способен проектировать, разрабатывать и поддерживать интегрированное программное обеспечение с использованием нейросетевых моделей и сквозных технологий искусственного интеллекта	ПК-2.2 Выбирает и моделирует архитектурные решения для реализации интегрированного программного обеспечения с использованием нейросетевых моделей и сквозных технологий искусственного интеллекта;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Обучение с подкреплением» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования. В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Обучение с подкреплением».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
УК-2	Способен управлять проектом на всех этапах его жизненного цикла	Технологическая (проектно-технологическая) практика (учебная); Правовые основы использования искусственного интеллекта; Прикладная статистика и анализ данных;	Управление проектами в сфере искусственного интеллекта;
ОПК-1	Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	Прикладная статистика и анализ данных; Глубокое обучение для NLP; Компьютерное зрение; SQL и NoSQL базы данных;	
ОПК-5	Способен устанавливать и сопровождать программное обеспечение информационных систем, осуществлять эффективное управление разработкой программных средств и проектов	Методы машинного обучения (продвинутый курс);	
ПК-2	Способен проектировать, разрабатывать и поддерживать интегрированное программное обеспечение с использованием нейросетевых моделей и сквозных технологий искусственного интеллекта	Методы машинного обучения (продвинутый курс); Компьютерное зрение; Глубокое обучение в компьютерном зрении; Программирование на языке C++ (продвинутый курс); Искусственный интеллект по отраслям**; Вайб-кодинг**;	Преддипломная практика; Генеративный искусственный интеллект; Искусственный интеллект в компьютерных играх**;

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Обучение с подкреплением» составляет «3» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			3
Контактная работа, ак.ч	36		36
Лекции (ЛК)	12		12
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	24		24
Самостоятельная работа обучающихся, ак.ч.	72		72
Контроль (экзамен/зачет с оценкой), ак.ч.	0		0
Общая трудоемкость дисциплины ак.ч.	ак.ч.	108	108
	зач.ед.	3	3

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы*

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
Раздел 1	Базовые принципы и классические алгоритмы RL	1.1	Основные понятия и постановка задачи RL: агент, среда, награды, цели обучения	Введение в обучение с подкреплением. Определение ключевых элементов: агент, среда, состояния, действия, награды. Постановка задачи максимизации накопленной награды. Различие между обучением с учителем, без учителя и обучением с подкреплением. Примеры применения RL. Формализация задачи через марковские процессы принятия решений (MDP).	ЛК
		1.2	Марковские процессы принятия решений (MDP), стратегии (policy), value-функции	Математическая формализация MDP: состояния, действия, функция переходов, функция награды, discount factor. Определение стратегии (policy) - детерминированной и стохастической. Value-функции: state-value функция $V(s)$ и action-value функция $Q(s,a)$. Уравнения Беллмана. Оптимальная стратегия и оптимальные value-функции.	ЛК
		1.3	Классические алгоритмы: Q-Learning, SARSA, исследование/эксплуатация, динамика обучения	Табличные методы обучения с подкреплением. Алгоритм Q-Learning как off-policy метод. SARSA как on-policy алгоритм. Exploration-exploitation trade-off и методы его решения (ϵ -greedy, softmax, UCB). Скорость сходимости, условия сходимости. Анализ динамики обучения и кривых learning curves.	ЛК
		1.4	Формализация задачи RL на примере специфического индустриального или исследовательского кейса	Практическое применение RL к конкретной задаче. Определение пространства состояний и действий. Разработка функции награды, учитывающей специфику задачи. Выбор appropriate алгоритма RL. Обсуждение вызовов реального мира: частичная наблюдаемость, задержки, шум в данных, ограничения безопасности.	СЗ
		1.5	Разработка и анализ простейшего агента RL в моделируемой среде (tabular RL)	Практическая реализация табличных методов RL. Создание простой среды (grid world, cart-pole). Имплементация Q-Learning или SARSA. Анализ результатов: сходимость, оптимальность стратегии. Визуализация обучения и полученной стратегии. Эксперименты с гиперпараметрами (learning rate, discount factor, exploration rate).	СЗ
		1.6	Разбор типичных ошибок и ограничения классических методов: instability, policy oscillation, погоня за наградой	Проблемы табличных методов при масштабировании. Нестабильность обучения и осцилляции в стратегии. Проблема reward hacking - эксплуатация артефактов функции награды. Катастрофическое забывание. Проклятие размерности. Необходимость функциональной аппроксимации для сложных задач.	СЗ
Раздел 2	Современные методы, deep RL и практические вызовы	2.1	Введение в deep RL: DQN, Policy Gradient, Actor-Critic, replay buffer	Использование глубоких нейронных сетей для аппроксимации value-функций и стратегий. Deep Q-Network (DQN): experience replay, target network, архитектура. Policy Gradient методы: REINFORCE, градиент стратегии. Actor-Critic архитектура: разделение на оценку value и выбор действий. Advantage функция.	ЛК
		2.2	Стабильность, robust regularization и explainability в глубоких RL-моделях	Проблемы стабильности обучения deep RL агентов. Методы регуляризации: entropy regularization, weight decay, градиентный clipping. Нормализация наблюдений и наград. Explainability в RL: визуализация внимания агента, saliency maps, интерпретация обученных представлений. Debugging deep RL агентов.	ЛК

		2.3	Индустриализация RL: автоматизация экспериментов, интеграция с внешними средами и	Инфраструктура для RL исследований: системы для отслеживания экспериментов (MLflow, Weights&Biases). Параллелизация обучения. Интеграция с реальными системами и симуляторами. Sim-to-real transfer. Этические аспекты: безопасность, fairness, прозрачность принятия решений, potential для злоупотреблений.	ЛК
--	--	-----	---	---	----

Номер раздела	Наименование раздела дисциплины	Наименование темы	Содержание темы	Вид учебной работы*
		этические вызовы		
		2.4 Построение и анализ агента deep RL в сложной среде (Atari/MuJoCo/OpenAI Gym)	Практическая реализация deep RL на benchmark задачах. Работа с Atari games или continuous control задачами в MuJoCo. Имплементация DQN, PPO или SAC. Preprocessing наблюдений. Tuning гиперпараметров. Оценка производительности, сравнение с baseline'ами. Анализ обученного поведения.	СЗ
		2.5 Практический разбор методов стабилизации и интерпретации deep RL: Grad-CAM, визуализация стратегий	Техники для понимания поведения deep RL агентов. Grad-CAM для визуализации, на что "смотрит" агент. Визуализация траекторий и стратегий в пространстве состояний. Ablation studies для понимания важности компонент. Анализ failure cases. Методы для диагностики проблем обучения.	СЗ
		2.6 Дискуссия: этика, безопасность, explainability и риски внедрения RL в реальные системы	Критическое обсуждение применения RL в production. Проблемы безопасности: непредсказуемое поведение, adversarial attacks. Safe RL: ограничения на действия, constrained optimization. Explainability требования в критических приложениях. Ответственность за решения агента. Социальные последствия автоматизации через RL. Регуляторные вызовы.	СЗ

* - заполняется только по ОЧНОЙ форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	
Семинарская	Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и техническими средствами мультимедиа презентаций.	
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Саттон, Р. С. Обучение с подкреплением: введение: практическое руководство / Р. С. Саттон, Э. Барто; пер. с англ. А. А. Слинкина. - Москва: ДМК Пресс, 2020. - 552 с. - ISBN 978-5-97060-097-9. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1210617>

2. Обучение с подкреплением для реальных задач: инженерный подход / Фил Уиндер; перевод с английского: Е. Черских. - Санкт-Петербург: БХВ-Петербург, 2023. - 400 с.: ил. ISBN 978-5-9775-6885-2

Дополнительная литература:

1. Жукова, Н. А. Обучение с подкреплением: лабораторные работы: учебно-методическое пособие / Н. А. Жукова, И. А. Куликов, А. Н. Субботин. — Санкт-Петербург: НИУ ИТМО, 2022. — 95 с. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/283880>

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров

- Электронно-библиотечная система РУДН – ЭБС РУДН <https://mega.rudn.ru/MegaPro/Web>
- ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>
- ЭБС «Юрайт» <http://www.biblio-online.ru>
- ЭБС «Консультант студента» www.studentlibrary.ru
- ЭБС «Знаниум» <https://znanium.ru/>

2. Базы данных и поисковые системы

- Sage <https://journals.sagepub.com/>
- Springer Nature Link <https://link.springer.com/>
- Wiley Journal Database <https://onlinelibrary.wiley.com/>
- Научометрическая база данных Lens.org <https://www.lens.org>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля*:

1. Курс лекций по дисциплине «Обучение с подкреплением».

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

РАЗРАБОТЧИКИ

Должность

РУКОВОДИТЕЛЬ БУП

Заведующий кафедрой

Должность

РУКОВОДИТЕЛЬ ОП ВО

Заведующий кафедрой прикладного
искусственного интеллекта

Должность

Подолько П.М.

Фамилия И.О

Подолько П.М.

Фамилия И.О

Подолько П.М.

Фамилия И.О
