МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Тверской государственный технический университет»
(ТвГТУ)

УТВЕРЖДАЮ
Проректор по учебной работе
_____ М.А.Смирнов
« _____ » _____ 20__ г.

РАБОЧАЯ ПРОГРАММА

Дисциплины, части формируемой участниками образовательных отношений Блока 1 «Дисциплины (модули)»
«Приложения систем обучения с подкреплением»

Направление подготовки бакалавров 09.03.04 Программная инженерия.

Направленность (профиль) — Разработка систем искусственного интеллекта.

Типы задач профессиональной деятельности — производственно-технологический.

Форма обучения – очная.

Факультет информационных технологий.

Кафедра «Программное обеспечение».

Тверь 20__

Рабочая программа дисциплины соответствует ОХОП подготовки бакалавров в части требований к результатам обучения по дисциплине и учебному плану.

Разработчик программы Е.И. Корнеева

Программа рассмотрена и одобрена на заседании кафедры ПО
«____» _____ 20_____ г., протокол № _____.

Заведующий кафедрой А. Л. Калабин

Согласовано
Начальник УМО Е.Э.Наумова

Начальник отдела
комплектования
зональной научной библиотеки О. Ф. Жмыхова

1. Цель и задачи дисциплины

Целью изучения дисциплины «Приложения систем обучения с подкреплением» является формирование у обучающихся представлений о методах и алгоритмах обучения с подкреплением, уяснение основных принципов разработки, внедрения и применения систем обучения с подкреплением для решения прикладных задач в области искусственного интеллекта.

Задачами дисциплины являются:

формирование представлений у обучающегося об основных понятиях и концепциях обучения с подкреплением;

формирование представлений об областях практического применения обучения с подкреплением и проблемах, связанных с его внедрением;

получение знаний о классических и современных алгоритмах обучения с подкреплением;

получение практических навыков разработки агентов обучения с подкреплением с использованием современных библиотек и фреймворков (OpenAI Gym, Stable Baselines3, RLlib).

2. Место дисциплины в структуре ОП

Дисциплина относится к части, формируемой участниками образовательных отношений Блока 1 ОП ВО, определяет подготовку бакалавров по направлению Программная инженерия в использовании методов искусственного интеллекта и обучения с подкреплением в дальнейшей учебной, научной и профессиональной деятельности. Она требует знания основ математического анализа, линейной алгебры, теории вероятностей и математической статистики, программирования на языке Python, основ машинного обучения.

В результате изучения данной дисциплины студенты будут знать основные методы обучения с подкреплением, уметь применять их для решения задач последовательного принятия решений с помощью средств современных библиотек, с учётом прикладной специфики.

3. Планируемые результаты обучения по дисциплине

3.1. Перечень компетенций, закреплённых за дисциплиной в ОХОП

Компетенция, закрепленная за дисциплиной в ОХОП:

ПК-5. *Способен разрабатывать, адаптировать, применять в профессиональной деятельности алгоритмы, программные средства, системы и комплексы обработки данных, методы и алгоритмы машинного обучения, программно-технические платформы, электронные библиотеки,*

программные оболочки приложений, сетевые технологии для решения задач в сфере искусственного интеллекта и смежных областях.

ПК-6. *Способен выбирать, применять и проводить экспериментальную проверку работоспособности программных компонентов систем, включающих модули по созданию искусственного интеллекта.*

Индикаторы компетенции, закрепленные за дисциплиной в ОХОП

ИПК-5.2. Проводит анализ решаемых задач и используемых алгоритмов, выявляет особенности часто используемых алгоритмов, предлагает показатели производительности алгоритмов при использовании приложений систем обучения с подкреплением в различных областях.

ИПК-5.3. Умеет работать с библиотеками и программными интерфейсами систем глубокого обучения и нейронных сетей.

ИПК-6.1. Выбирает, комбинирует и адаптирует существующие программные продукты, для решения необходимых функций, профессиональных задач предприятий или организаций.

ИПК-6.2. Самостоятельно создает на основе стандартных оболочек с привлечением искусственного интеллекта программное обеспечение для решения необходимых функций, профессиональных задач предприятий или организаций.

Показатели оценивания индикаторов достижения компетенций

Знать:

З1. Основные концепции обучения с подкреплением: агент, среда, состояние, действие, награда, политика.

З2. Классические алгоритмы обучения с подкреплением: Q-learning, SARSA, Policy Gradient.

З3. Современные алгоритмы глубокого обучения с подкреплением: DQN, A3C, PPO, DDPG, SAC.

З4. Области применения обучения с подкреплением: игры, робототехника, рекомендательные системы, автономное управление.

Уметь:

У1. Формализовать прикладные задачи в терминах обучения с подкреплением (MDP, POMDP).

У2. Реализовывать агенты обучения с подкреплением с использованием библиотек Python (OpenAI Gym, Stable Baselines3).

У3. Оценивать эффективность алгоритмов обучения с подкреплением и проводить их сравнительный анализ.

Иметь опыт практической подготовки:

ПП1. работы с современными инструментами и технологиями для обработки больших данных.

3.2. Технологии, обеспечивающие формирование компетенций

Проведение лекционных и практических занятий, выполнение курсовой работы, самостоятельная работа под руководством преподавателя.

4. Трудоёмкость дисциплины и виды учебной работы

Таблица 1. Распределение трудоемкости дисциплины по видам учебной работы

Вид учебной работы	**Зачётных единиц**	**Академических часов**
Общая трудоемкость дисциплины	**2**	**72**
Аудиторные занятия (всего)		45
В том числе:		
Лекции		15
Практические занятия (ПЗ)		не предусмотрены
Лабораторные работы (ЛР)		30
Самостоятельная работа (всего)		**27**
В том числе:		
Курсовая работа (КР)		не предусмотрена
Курсовой проект (КП)		не предусмотрен
Расчетно-графические работы		не предусмотрены
Другие виды самостоятельной работы:		
- подготовка к защите лабораторных работ		18
- изучение теоретического материала		9
Контроль текущий и промежуточный (балльно-рейтинговый, зачёт)		0
Практическая подготовка при реализации дисциплины (всего)		30
В том числе:		
Практические занятия (ПЗ)		не предусмотрены
Лабораторные работы (ЛР)		30

Курсовая работа (КП)		не предусмотрена
Курсовой проект (КР)		не предусмотрен

5. Содержание дисциплины

5.1. Структура дисциплины

Таблица 2. Структура дисциплины

№	Наименование модуля	Труд-ть часы	Лекции	Практич. занятия	Лаб. работы	Сам. работы
1	Введение в обучение с подкреплением	14	3	-	6	5
2	Марковские процессы принятия решений	12	2	-	6	4
3	Алгоритмы обучения на основе значений	14	3	-	6	5
4	Методы градиента политики	14	3	-	6	5
5	Глубокое обучение с подкреплением и приложения	18	4	-	6	8
Всего на дисциплину		**72**	**15**		**30**	**27**

5.2. Содержание разделов (модулей, блоков) дисциплины

МОДУЛЬ 1. «Введение в обучение с подкреплением»

Основные концепции обучения с подкреплением. Агент, среда, состояния, действия, награды. Отличия от обучения с учителем и обучения

без учителя. История развития обучения с подкреплением. Примеры применения: игры (шахматы, Go, компьютерные игры), робототехника, рекомендательные системы, автономные транспортные средства, оптимизация ресурсов. Установка и настройка OpenAI Gym. Создание и использование простых сред.

МОДУЛЬ 2. «Марковские процессы принятия решений»

Марковский процесс принятия решений (MDP): определение, основные компоненты. Марковское свойство. Политика, функция ценности состояния, функция ценности действия. Уравнения Беллмана. Оптимальная политика и оптимальные функции ценности. Методы динамического программирования: оценка политики (Policy Evaluation), улучшение политики (Policy Improvement), итерация по политике (Policy Iteration), итерация по значению (Value Iteration). Частично наблюдаемые МПР (POMDP).

МОДУЛЬ 3. «Алгоритмы обучения на основе значений»

Методы Монте-Карло: оценка политики методом Монте-Карло, управление методом Монте-Карло с ε-жадной стратегией. Временные различия (Temporal Difference, TD): TD(0), сравнение с методами Монте-Карло. Q-learning: принцип работы, уравнение обновления, сходимость. SARSA (State-Action-Reward-State-Action): on-policy алгоритм, отличия от Q-learning. Исследование vs эксплуатация (exploration vs exploitation): ε-жадная стратегия, Softmax, UCB (Upper Confidence Bound). Deep Q-Networks (DQN): архитектура, Experience Replay, Target Network, Double DQN, Dueling DQN.

МОДУЛЬ 4. «Методы градиента политики»

Основы методов градиента политики. Теорема градиента политики (Policy Gradient Theorem). REINFORCE: алгоритм, базовая линия (baseline). Actor-Critic методы: комбинация оценки ценности и градиента политики. Advantage Actor-Critic (A2C). Asynchronous Advantage Actor-Critic (A3C).

Proximal Policy Optimization (PPO): Trust Region Policy Optimization (TRPO), упрощенная реализация PPO. Deep Deterministic Policy Gradient (DDPG): применение для непрерывных пространств действий. Twin Delayed DDPG (TD3). Soft Actor-Critic (SAC).

МОДУЛЬ 5. «Глубокое обучение с подкреплением и приложения»

Современные архитектуры нейронных сетей для RL: CNN, RNN, LSTM, Transformer. Мультиагентное обучение с подкреплением (MARL): координация агентов, соревновательные и кооперативные сценарии. Обратное обучение с подкреплением (Inverse Reinforcement Learning). Имитационное обучение (Imitation Learning). Применение обучения с подкреплением: игры (Atari, Dota 2, StarCraft II), робототехника (манипуляция объектами, навигация), рекомендательные системы, управление трафиком, финансы, энергетика. Фреймворки для RL: Stable Baselines3, RLlib, TF-Agents. Практические аспекты: отладка, подбор гиперпараметров, мониторинг обучения, оценка производительности агентов.

5.3. Лабораторные работы

Таблица 3. Лабораторные работы и их трудоемкость

Модули. Цели лабораторных работ	**Примерная тематика лабораторных работ**	**Трудоемкость в часах**
Модуль 1 **Цель:** освоение основ обучения с подкреплением, работа с OpenAI Gym, реализация простых агентов	Знакомство с OpenAI Gym. Реализация случайного агента для простых сред (CartPole, MountainCar)	6
Модуль 2 **Цель:** изучение методов динамического программирования и их применение для решения задач принятия решений	Реализация методов динамического программирования: Policy Iteration и Value Iteration для дискретной среды	6
Модуль 3 **Цель:** освоение	Реализация алгоритмов Q-learning и SARSA.	6

алгоритмов обучения на основе значений, включая Q-learning, SARSA и Deep Q-Networks	Сравнение on-policy и off-policy подходов	
	Реализация Deep Q-Network (DQN) с использованием PyTorch или TensorFlow для среды Atari	6
Модуль 4-5 **Цель:** изучение методов градиента политики и современных алгоритмов глубокого обучения с подкреплением	Реализация алгоритма Proximal Policy Optimization (PPO) с использованием Stable Baselines3. Применение для задачи управления непрерывными действиями	6

5.4. Практические занятия.

Учебным планом практические занятия не предусмотрены.

6.Самостоятельная работа обучающихся и текущий контроль успеваемости.

6.1. Цели самостоятельной работы

Формирование способностей к самостоятельному познанию и обучению, поиску литературы, обобщению, оформлению и представлению полученных результатов, их критическому анализу, поиску новых и неординарных решений, аргументированному отстаиванию своих предложений, умений подготовки выступлений и ведения дискуссий.

6.2. Организация и содержания самостоятельной работы

Самостоятельная работа заключается в изучении отдельных тем курса по заданию преподавателя по рекомендуемой им учебной литературе, в решении упражнений, в подготовке к практическим занятиям, к текущему контролю успеваемости, зачету, в выполнении курсовой работы. После вводных практических занятий, в которых обозначается содержание дисциплины, ее проблематика и практическая значимость, студентам

выдаются темы курсовой работы, определяется порядок подготовки доклада и презентации для ее защиты.

Текущий контроль успеваемости осуществляется с использованием модульно-рейтинговой системы обучения и оценки текущей успеваемости обучающихся в соответствии с СТО СМК 02.102-2012.

7. Учебно-методическое и информационное обеспечение дисциплины

7.1. Основная литература по дисциплине

1. Горюшкин, А. П. Математическая логика и теория алгоритмов : учебник / А. П. Горюшкин. — Саратов : Вузовское образование, 2022. — 499 c. — ISBN 978-5-4487-0808-4. — Текст : электронный // IPR SMART : [сайт]. — URL: https://www.iprbookshop.ru/117296.html . - (ID=144996-0)

2. Крупский, В. Н. Теория алгоритмов. Введение в сложность вычислений : учебное пособие для вузов / В. Н. Крупский. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2022. — 117 с. — (Высшее образование). — ISBN 978-5-534-04817-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/492937 . - (ID=142651-0).

3. Баланов, А. Н. Искусственный интеллект. Понимание, применение и перспективы : учебник для вузов / А. Н. Баланов. — 2-е изд., стер. — Санкт-Петербург : Лань, 2025. — 312 с. — ISBN 978-5-507-52357-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/448697 (дата обращения: 15.12.2025). — Режим доступа: для авториз. пользователей. - (ID=161671-0)

7.2. Дополнительная литература по дисциплине

1. Вайнштейн, Ю. В. Математическая логика и теория алгоритмов : учебное пособие / Ю. В. Вайнштейн, Т. Г. Пенькова, В. И. Вайнштейн. — Красноярск : СФУ, 2019. — 110 с. — ISBN 978-5-7638-4076-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/157585 . - (ID=145006-0)

2. Виноградов, Г.П. Теория алгоритмов и элементы теории формальных языков : учеб. пособие для студентов направлений подготовки бакалавра 15.03.04 Автоматизация технол. процессов и производств профиля "Технология и автоматизация производства в машиностроении" и 09.03.02 Информ. системы и технологии : в составе учебно-методического комплекса / Г.П. Виноградов, В.Н. Богатиков; Тверской государственный технический университет, Кафедра ИПМ. -

Тверь :ТвГТУ, 2016. - (УМК-У). - Сервер. - Текст : электронный. - ISBN 978-5-7995-0845-6 : 0-00. - URL: http://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/113354 . - (ID=113354-1)

3. Гамова, А.Н. Математическая логика и теория алгоритмов : учебное пособие / А.Н. Гамова. - 4-е изд., доп. - Саратов : Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского, 2020. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.09.2022. - ЭБС Лань. - ISBN 978-5-292-04649-3. - URL: https://e.lanbook.com/book/170590 . - (ID=111573-0)

4. Глухов, М.М. Математическая логика. Дискретные функции. Теория алгоритмов : учебное пособие для вузов по спец. и напр. по информ. безопасности / М.М. Глухов, А.Б. Шишков. - Санкт-Петербург [и др.] : Лань, 2012. - 405 с. - (Учебники для вузов. Специальная литература). - Текст : непосредственный. - ISBN 978-5-8114-1344-7 : 766 р. 92 к. - (ID=95689-3)

5. Задачи и упражнения по математической логике, дискретным функциям и теории алгоритмов / М. М. Глухов, О. А. Козлитин, В. А. Шапошников, А. Б. Шишков. — 3-е изд., стер. — Санкт-Петербург : Лань, 2022. — 112 с. — ISBN 978-5-507-44852-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/247400 (дата обращения: 15.12.2025). — Режим доступа: для авториз. пользователей. - - (ID=189488-0)

6. Игошин, В.И.З адачи и упражнения по математической логике и теории алгоритмов : учеб. пособие для вузов по спец. 032100 "Математика" / В.И. Игошин. - 4-е изд. - М. : Академия, 2008. - 303 с. - (Высшее профессиональное образование). - Текст : непосредственный. - ISBN 978-5- 7695-5272-4 : 333 р. 30 к. - (ID=87399-15)

7. Лавров, И.А.Задачи по теории множеств, математической логике и теории алгоритмов : [учеб. пособие для вузов] / И.А. Лавров, Л.Л. Максимова. - 5-е изд. ;испр. - Москва :Физматлит, 2004. - 255 с. - Библиогр. : с. 248 - 249. - Текст : непосредственный. - ISBN 5-9221-0026-2 : 134 р. 64 к. - (ID=22585-5)

8. Судоплатов, С. В. Математическая логика и теория алгоритмов : учебник и практикум для вузов / С. В. Судоплатов, Е. В. Овчинникова. — 5-е изд., стер. — Москва : Издательство Юрайт, 2025. — 207 с. — (Высшее образование). — ISBN 978-5-534-12274-9. — Текст : электронный //

Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/559978 - (ID=142652-0)

9. Широков, Д. В. Теория алгоритмов : учебное пособие / Д. В. Широков. — Киров :ВятГУ, 2017. — 163 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/134610 . - (ID=111523-0).

7.3. Методические материалы

1. Учебно-методический комплекс дисциплины обязательной части Блока 1 "Приложения систем обучения с подкреплением". Направление подготовки 09.03.04 Программная инженерия. Направленность (профиль) - Разработка систем искусственного интеллекта : ФГОС 3++ / Каф. Программное обеспечение ; сост. Е.И. Корнеева. - 2025. - (УМК). - Текст : электронный. - URL: https://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/189487 . - (ID=189487-0)

7.4. Программное обеспечение

Операционная система Microsoft Windows: лицензии № ICM-176609 и № ICM-176613 (Azure Dev Tools for Teaching).

Microsoft Office 2007 Russian Academic: OPEN No Level: лицензия № 41902814.

7.5. Специализированные базы данных, справочные системы, электронно-библиотечные системы, профессиональные порталы в Интернет

ЭБС и лицензионные ресурсы ТвГТУ размещены:

ЭБС и лицензионные ресурсы ТвГТУ размещены:

1. Ресурсы: https://lib.tstu.tver.ru/header/obr-res
2. ЭБ ТвГТУ: https://elib.tstu.tver.ru/MegaPro/Web
3. ЭБС "Лань": https://e.lanbook.com/
4. ЭБС "Университетская библиотека онлайн": https://biblioclub.ru/
5. Национальная электронная библиотека: https://rusneb.ru
6. ЦОР IPRSmart: https://www.iprbookshop.ru/
7. Электронная образовательная платформа "Юрайт": https://urait.ru/
8. Научная электронная библиотека eLIBRARY: https://elibrary.ru/
9. Информационная система "ТЕХНОРМАТИВ". Конфигурация "МАКСИМУМ" : сетевая версия (годовое обновление) : [нормативно-технические, нормативно-правовые и руководящие документы (ГОСТы, РД, СНиПы и др.]. Диск 1, 2, 3, 4. - М. :Технорматив, 2014. -

(Документация для профессионалов). - CD. - Текст : электронный. - 119600 р. – (105501-1)

10. База данных учебно-методических комплексов: https://lib.tstu.tver.ru/header/umk.html

УМК размещен: https://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/189487

8. Материально-техническое обеспечение дисциплины

При изучении дисциплины «Приложения систем обучения с подкреплением» используются современные средства обучения: наглядные пособия, диаграммы, схемы.

Возможна демонстрация лекционного материала с помощью оверхедпроектора (кодоскопа) и мультипроектора.

Вуз имеет лабораторию для реализации лабораторного практикума; учебный класс для проведения самостоятельной работы по курсу «Приложения систем обучения с подкреплением», оснащенный современной компьютерной и офисной техникой, необходимым программным обеспечением, электронными учебными пособиями, имеющий безлимитный выход в глобальную сеть; аудиторию для проведения презентаций студенческих работ, оснащенную аудиовизуальной техникой.

9. Оценочные средства для проведения промежуточной аттестации

9.1. Оценочные средства для проведения промежуточной аттестации в форме экзамена

Учебным планом экзамен по дисциплине не предусмотрен.

9.2. Оценочные средства для проведения промежуточной аттестации в форме зачета

1. Шкала оценивания промежуточной аттестации – «зачтено», «не зачтено».

2. Вид промежуточной аттестации в форме зачёта.

Вид промежуточной аттестации устанавливается преподавателем: по результатам текущего контроля знаний обучающегося без дополнительных

контрольных испытаний или с выполнением дополнительного итогового контрольного испытания при наличии задолженностей в текущем контроле.

4. Для дополнительного итогового контрольного испытания студенту в обязательном порядке предоставляется:

база заданий, предназначенных для предъявления обучающемуся на дополнительном итоговом контрольном испытании (типовой образец задания приведен в Приложении), задание выполняется письменно;

методические материалы, определяющие процедуру проведения дополнительного итогового испытания и проставления зачёта.

Число заданий для дополнительного итогового контрольного испытания – 20.

Число вопросов – 3 (1 вопрос для категории «знать» и 2 вопроса для категории «уметь»).

Перечень вопросов для дополнительного итогового испытания:

1. Основные компоненты системы обучения с подкреплением: агент, среда, состояние, действие, награда.
2. Определение Марковского процесса принятия решений (MDP). Марковское свойство.
3. Политика в обучении с подкреплением: детерминированная и стохастическая политика.
4. Функция ценности состояния V(s) и функция ценности действия Q(s,a).
5. Уравнения Беллмана для функции ценности состояния и действия.
6. Оптимальная политика и оптимальные функции ценности.
7. Методы динамического программирования: Policy Evaluation, Policy Improvement.
8. Алгоритмы Policy Iteration и Value Iteration: принципы работы и отличия.
9. Частично наблюдаемые марковские процессы принятия решений (POMDP).
10. Методы Монте-Карло в обучении с подкреплением: особенности и применение.

11. Временные различия (Temporal Difference, TD): основные идеи и преимущества.
12. Алгоритм TD(0): принцип работы и уравнение обновления.
13. Алгоритм Q-learning: принцип работы, уравнение обновления Q-значений.
14. Алгоритм SARSA: отличия от Q-learning, on-policy vs off-policy.
15. Проблема исследования vs эксплуатации (exploration vs exploitation).
16. ε-жадная стратегия выбора действий.
17. Softmax и UCB (Upper Confidence Bound) стратегии.
18. Аппроксимация функций ценности с помощью нейронных сетей.
19. Deep Q-Network (DQN): архитектура и основные компоненты.
20. Experience Replay в DQN: назначение и принцип работы.
21. Target Network в DQN: зачем нужна и как работает.
22. Double DQN: решение проблемы переоценки Q-значений.
23. Dueling DQN: архитектура и преимущества.
24. Методы градиента политики (Policy Gradient): основные идеи.
25. Теорема градиента политики (Policy Gradient Theorem).
26. Алгоритм REINFORCE: принцип работы и особенности.
27. Базовая линия (baseline) в методах градиента политики.
28. Actor-Critic методы: комбинация оценки ценности и градиента политики.
29. Advantage Actor-Critic (A2C) и Asynchronous A3C.
30. Proximal Policy Optimization (PPO): основные идеи и преимущества.
31. Trust Region Policy Optimization (TRPO).
32. Deep Deterministic Policy Gradient (DDPG) для непрерывных пространств действий.
33. Twin Delayed DDPG (TD3): улучшения над DDPG.
34. Soft Actor-Critic (SAC): максимизация энтропии в RL.
35. OpenAI Gym: основные компоненты и использование.

36. Создание и регистрация пользовательских сред в OpenAI Gym.
37. Stable Baselines3: основные возможности библиотеки.
38. Обучение агента с использованием Stable Baselines3.
39. Оценка и тестирование обученного агента.
40. Визуализация процесса обучения агента.
41. Подбор гиперпараметров в алгоритмах обучения с подкреплением.
42. Проблема кредитного назначения (credit assignment problem).
43. Discount factor (γ): влияние на поведение агента.
44. Reward shaping: модификация функции награды.
45. Применение обучения с подкреплением в играх.
46. Применение обучения с подкреплением в робототехнике.
47. Применение обучения с подкреплением в рекомендательных системах.
48. Применение обучения с подкреплением в автономном управлении.
49. Мультиагентное обучение с подкреплением (MARL).
50. Обратное обучение с подкреплением (Inverse Reinforcement Learning).
51. Имитационное обучение (Imitation Learning).
52. Model-based vs model-free методы в RL.
53. Планирование в обучении с подкреплением.
54. Иерархическое обучение с подкреплением.
55. Transfer learning в обучении с подкреплением.
56. Meta-learning в контексте RL.
57. Проблемы стабильности обучения в deep RL.
58. Проблема sample efficiency в обучении с подкреплением.
59. Offline reinforcement learning.
60. Этические аспекты применения обучения с подкреплением.

Критерии выполнения контрольного испытания и условия проставления зачета:

Критерии оценки и её значение для категории “знать” (бинарный критерий):

ниже базового - 0 балл;

базовый уровень – 1 балла;

Критерии оценки и её значение для категории “уметь” (бинарный критерий):

отсутствие умения – 0 баллов;

наличие умения – 2 балла.

Критерии итоговой оценки за зачет:

«зачтено» – при сумме баллов 3, 4 или 5;

«не зачтено» – при сумме баллов 0, 1 или 2.

10.Методические рекомендации по организации изучения дисциплины

Студенты перед началом изучения дисциплины ознакомлены с системами кредитных единиц и балльно-рейтинговой оценки, которые должны быть опубликованы и размещены на сайте вуза или кафедры.

Студенты, изучающие дисциплину обеспечиваются электронными изданиями или доступом к ним, учебно-методическим комплексом по дисциплине.

11.Внесение изменений и дополнений в рабочую программу дисциплины

Кафедра ежегодно обновляет содержание рабочих программ дисциплин, которые оформляются протоколами. Форма протокола утверждена Положением о структуре, содержании и оформлении рабочих программ дисциплин, по образовательным программ, соответствующих ФГОС ВО с учетом профессиональных стандартов.

Приложение 1

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Тверской государственный технический университет»

Направление подготовки бакалавров 09.03.04 Программная инженерия.

Направленность (профиль) — Разработка систем искусственного интеллекта.

Кафедра «Программного обеспечения»

Дисциплина «Приложения систем обучения с подкреплением»

ЗАДАНИЕ ДЛЯ ДОПОЛНИТЕЛЬНОГО ИТОГОВОГО КОНТРОЛЬНОГО ИСПЫТАНИЯ № 1

1. Вопрос для проверки уровня «ЗНАТЬ» по разделу «Основы обучения с подкреплением» – 0, 1 или 2 балла

Дайте определения следующих понятий: Марковский процесс принятия решений (MDP), марковское свойство, политика (детерминированная и стохастическая), функция ценности состояния V(s) и функция ценности действия Q(s,a), уравнения Беллмана.

2. Задание для проверки уровня «УМЕТЬ» – 0 или 1 балл:

Опишите алгоритм Q-learning: принцип работы, уравнение обновления Q-значений, отличие от SARSA (on-policy vs off-policy). Объясните, как решается проблема exploration vs exploitation с помощью ε-жадной стратегии.

3. Задание для проверки уровня «УМЕТЬ» – 0 или 2 балла.

Объясните архитектуру Deep Q-Network (DQN) и роль ключевых компонентов: Experience Replay и Target Network. Докажите, почему без этих компонентов обучение DQN нестабильно. Опишите, как реализовать обучение агента DQN с использованием библиотеки Stable Baselines3 для среды CartPole-v1 из OpenAI Gym.

Критерии итоговой оценки за зачет:

«зачтено» – при сумме баллов 3, 4 или 5;
«не зачтено» – при сумме баллов 0, 1 или 2.

Составитель: _____ Е.И. Корнеева

Заведующий кафедрой ПО:

д.ф.-м.н., профессор _____ Калабин А.Л.