

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ І.І. МЕЧНИКОВА
Кафедра математичного забезпечення комп'ютерних систем



“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної роботи

[Signature]

20 25 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ВК1 «Введення в навчання з підкріпленням»

(назва навчальної дисципліни)

Рівень вищої освіти _____ Другий (магістерський) _____

Галузь знань _____ F – Інформаційні технології _____

Спеціальність _____ F7 – Комп'ютерна інженерія _____
(код і назва спеціальності (тей))

Освітньо-професійна програма _____ Комп'ютерна інженерія _____
(назва ОПП)

Робоча програма навчальної дисципліни
«Введення в навчання з підкріпленням»: – Одеса: ОНУ, 2025. – 14 с.

Розробник:
Пенко В.Г., к.т.н., доцент кафедри МЗКС

Робоча програма затверджена на засіданні кафедри математичного забезпечення комп'ютерних систем

Протокол № 1 від. " 10 " 09 2025

Завідувач кафедри


(підпис)

(Євгеній МАЛАХОВ)
(Ім'я ПРІЗВИЩЕ)

Погоджено із гарантом ОПП «Комп'ютерна інженерія»


(підпис)

(Юрій ГУНЧЕНКО)
(Ім'я ПРІЗВИЩЕ)

Схвалено навчально-методичною комісією (НМК) з ІТ факультету МФІТ

Протокол № 2 від. " 10 " 09 2025 р.

Голова НМК


(підпис)

(Лариса МАРТИНОВИЧ)
(Ім'я ПРІЗВИЩЕ)

Переглянуто та затверджено на засіданні кафедри _____

Протокол № ____ від. " ____ " _____ 20__ р.

Завідувач кафедри

(підпис)

(_____)
(Ім'я ПРІЗВИЩЕ)

Переглянуто та затверджено на засіданні кафедри _____

Протокол № ____ від. " ____ " _____ 20__ р.

Завідувач кафедри

(підпис)

(_____)
(Ім'я ПРІЗВИЩЕ)

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни	
		денна форма навчання	заочна форма навчання
Загальна кількість: кредитів – 3 годин – 90 змістових модулів – 2	Галузь знань <u>F – Інформаційні технології</u> (шифр і назва)	<i>Вибіркова</i>	
		<i>Рік підготовки:</i>	
	Спеціальність <u>F7 – Комп’ютерна інженерія</u> (шифр і назва)	1	
		<i>Семестр</i>	
		2	
		<i>Лекції</i>	
	Рівень вищої освіти: <u>Другий</u> (<u>магістерський</u>)	16 год.	
		<i>Практичні, семінарські</i>	
		14 год	
		<i>Лабораторні</i>	
		год.	год
		<i>Самостійна робота</i>	
		60 год.	год
		Форма підсумкового контролю: залік	

* – за наявності

2. Мета та завдання навчальної дисципліни

Метою курсу є:

- вивчення основних підходів для вирішення базових завдань навчання з підкріплення.
- набуття практичних навичок виконання обчислювальних експериментів, що демонструють теоретичний матеріал курсу.

Завдання:

- сформулювати розуміння ролі навчання з підкріпленням (Reinforcement Learning, RL) у сучасній системі методів штучного інтелекту та машинного навчання;
- ознайомити студентів із математичним апаратом RL, зокрема з поняттями станів, дій, винагород і Марковських процесів прийняття рішень (MDP);
- розвинути здатність аналізувати задачі прийняття рішень у стохастичних середовищах, визначати політику поведінки агента та оцінювати її ефективність;
- навчити застосовувати базові алгоритми RL — динамічне програмування, ітерацію по цінностях, Монте-Карло, TD-навчання, SARSA та Q-learning — у моделюванні поведінки агентів;
- сформулювати практичні навички використання програмних інтерфейсів типу OpenAI Gym для моделювання взаємодії агентів із середовищем;
- розвинути вміння проектувати та тестувати навчальні середовища, аналізувати збіжність і стабільність алгоритмів навчання;
- показати зв'язок між класичними алгоритмами RL та сучасними нейромережевими підходами (глибоке навчання, Deep Q-Network);
- сприяти формуванню системного мислення щодо побудови автономних агентів, здатних до адаптивного прийняття рішень у невизначених умовах;
- підготувати студентів до подальшого вивчення просунутих тем, таких як градієнти політики, actor-critic моделі, багатоагентне навчання та прикладні застосування RL;
- розвинути здатність критично оцінювати сучасні наукові розробки в галузі RL та адаптувати їх до власних дослідницьких або інженерних завдань;

Процес вивчення дисципліни спрямований на формування елементів наступних **компетентностей:**

а) загальних

ЗК3. Здатність проводити дослідження на відповідному рівні.

б) спеціальних (фахових):

СК7. Здатність досліджувати, проектувати та моделювати елементи кіберфізичних систем з використанням сучасних науково-технічних методів

СК11. Здатність обирати ефективні методи розв'язування складних задач комп'ютерної інженерії, критично оцінювати отримані результати та аргументувати прийняті рішення.

Програмні результати навчання:

РН1. Застосовувати загальні підходи пізнання, методи математики, природничих та інженерних наук до розв'язання складних задач комп'ютерної інженерії.

РН2. Знаходити необхідні дані, аналізувати та оцінювати їх.

РН3. Будувати та досліджувати моделі комп'ютерних систем і мереж, оцінювати їх адекватність, визначати межі застосовності.

РН6. Аналізувати проблематику, ідентифікувати та формулювати конкретні проблеми, що потребують вирішення, обирати ефективні методи їх вирішення.

РН7. Вирішувати задачі аналізу та синтезу комп'ютерних систем та мереж.

РН14. Планувати і виконувати наукові дослідження в сфері комп'ютерної інженерії, формулювати і перевіряти гіпотези, обирати методики та інструменти, аналізувати результати, обґрунтовувати висновки.

РН15. Застосовувати сучасні науково-технічні методи досліджування, проектування та моделювання елементів кіберфізичних систем.

Очікувані результати навчання. У результаті вивчення навчальної дисципліни студент повинен:

знати: основні концепції навчання з підкріпленням та способи програмного експериментування для досягнення цілей навчання з підкріпленням.

вміти: будувати математичні моделі базових типів завдань навчання з підкріпленням, розробляти програмне забезпечення, що реалізує базові алгоритми навчання з підкріпленням.

3. Зміст навчальної дисципліни

Змістовний модуль 1 Основні концепції навчання з підкріпленням.

Тема 1. Місце навчання з підкріпленням в сучасному баченні штучного інтелекту.

- Співставлення концепцій штучного інтелекту та машинного навчання;
- Різновиди машинного навчання.

Література: [1, 2, 4].

Тема 2. Інтерфейс GYM як стандарт представлення середовища в контексті навчання з підкріпленням.

- Опис елементів інтерфейсу;
- Експериментування з середовищами різних типів.

Література: [8].

Тема 3 Марковський процес прийняття рішення.

- Інтерфейс між агентом та навколишнім середовищем
- Цілі та винагороди
- Дохід та епізоди
- Уніфікована нотація для епізодичних та безперервних завдань
- Стратегії та функції цінності
- Оптимальні стратегії та оптимальні функції цінності

Література: [1, 2].

Тема 4 Динамічне програмування.

- Представлення МППР у вигляді графу станів;

- Рівняння Беллмана;
- Застосування динамічного програмування для рівняння Беллмана;
- Асинхронне динамічне програмування;
- Ефективність динамічного програмування.

Література: [1, 2, 6, 8].

Тема 5 Алгоритми ітерації по цінностям та по стратегії.

- Оцінювання стратегії;
- Поліпшення стратегії;
- Ітерація зі стратегій;
- Ітерація за цінністю;
- Узагальнена ітерація зі стратегій.

Література: [1, 2, 6, 8].

Змістовний модуль 2 Безмодельні алгоритми навчання з підкріпленням.

Тема 6. Методи Монте-Карло. Оцінювання стратегії та проблема дослідження.

- Прогноз методами Монте-Карло;
- Оцінювання цінності дій методом Монте-Карло;
- Управління методом Монте-Карло;
- Управління методом Монте-Карло без дослідницьких стартів;
- Прогноз з розділеною стратегією у вигляді вибірки за значимістю;
- Інкрементна реалізація;
- Управління методом Монте-Карло із розділеною стратегією.

Література: [6, 10].

Тема 7. TD-навчання.

- Прогноз TD-методами;
- Переваги TD-методів передбачення;
- Оптимальність TD(0).

Література: [1, 6].

Тема 8. Алгоритм SARSA.

- Sarsa: TD-управління з єдиною стратегією;
- Q-навчання: TD-управління з розділеною стратегією;
- Expected Sarsa;
- Зміщення максимізації та подвійне навчання.

Література: [1, 4, 7, 9].

Тема 9. Глибока Q-мережа. Online Q-навчання.

- Від табличного Q-навчання до глибокого Q-навчання;
- Нейронно-орієнтована Q-ітерація;
- Практичні міркування та труднощі щодо нейронно-орієнтованої Q-ітерації;
- Онлайн Q-навчання.

-

Література: [1, 3].

Тема 10. Алгоритм DQN.

- Смертельна тріада;
- Повторення досвіду;
- Цільова мережа;

- Функція втрат;
 - Алгоритм DQN.
- Література: [1, 3, 6].

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин									
	Денна форма					Заочна форма				
	Усього	у тому числі				Усього	у тому числі			
		л	п	лаб	ср		л	п	лаб	ср
1	2	3	4	5	6	7	8	9	10	11
Змістовний модуль 1. Основні концепції навчання з підкріпленням										
Тема 1. Місце навчання з підкріпленням в сучасному баченні штучного інтелекту.	7	1			6					
Тема 2. Інтерфейс GYM як стандарт представлення середовища в контексті навчання з підкріпленням.	9	1	2		6					
Тема 3. Марковський процес прийняття рішення.	8	1	1		6					
Тема 4. Динамічне програмування.	8	1	1		6					
Тема 5. Алгоритми ітерації по цінностям та по стратегії.	9	2	1		6					
Змістовний модуль 2. Безмодельні алгоритми навчання з підкріпленням										
Тема 6. Методи Монте-Карло. Оцінювання стратегії та проблема дослідження.	9	2	1		6					
Тема 7. TD-навчання.	10	2	2		6					
Тема 8. Алгоритм SARSA.	10	2	2		6					
Тема 9. Глибока Q-мережа. Online Q-навчання.	10	2	2		6					
Тема 10. Алгоритм DQN.	10	2	2		6					
Всього годин	90	16	14		60					

5. Теми семінарських занять

Семінарські заняття не передбачені

6. Теми практичних занять

№ з/п	Назва теми	Кількість годин	
		денне	заочне
1	Виконання демонстраційних обчислювальних експериментів з декількома стандартними середовищами.	2	
2	Побудова простих прикладів МППР та обчислення на графі станів.	1	
3	Обчислення цінності станів та дій рівнянням Беллмана.	1	
4	Реалізація алгоритмів ітерації по цінностям та по стратегії.	1	
5	Реалізація оцінювання стратегії методом Монте-Карло з використанням епсілон-жадібною стратегії.	1	
6	Реалізація алгоритма TD-навчання.	2	
7	Реалізація алгоритма SARSA.	2	
8	Реалізація алгоритма Neuro-fitted iteration	2	
9	Реалізація алгоритма DQN.	2	
	Разом	14	

7. Теми лабораторних занять

Лабораторні заняття не передбачені

8. Самостійна робота

№ з/п	Назва теми	Кількість годин	
		денне	заочне
1	Місце навчання з підкріпленням в сучасному баченні штучного інтелекту.	6	
2	Інтерфейс GYM як стандарт представлення середовища в контексті навчання з підкріпленням.	6	
3	Марковський процес прийняття рішення.	6	
4	Динамічне програмування.	6	
5	Алгоритми ітерації по цінностям та по стратегії.	6	
6	Методи Монте-Карло. Оцінювання стратегії та проблема дослідження.	6	
7	TD-навчання.	6	
8	Алгоритм SARSA.	6	
9	Глибока Q-мережа. Online Q-навчання.	6	
10	Алгоритм DQN.	6	
	Разом	60	

До самостійної роботи відноситься:

[1] – підготовка до лекцій та лабораторних занять;

В якості самостійної роботи здобувач може пройти професійні курси, тренінги, професійне стажування. онлайн-курси. громадську освіту, отримати відповідний сертифікат на освітніх платформах, що може бути зараховано як періодичний, поточний та фрагмент підсумкового контролю. Пропонуються до розгляду такі курси, які можуть бути визнані як періодичний, поточний контроль та/або бонусні бали (за бажанням здобувача):

Назва освітньої платформи, назва онлайн курсу, короткий опис, посилання	Дотично до теми (та/або компетентності, результати навчання, що формуються)
Prometheus. Машинне навчання— Безкоштовний курс з основ машинного навчання, включає розділ «навчання з підкріпленням» як частину програми. (https://prometheus.org.ua/prometheus-free/machine-learning/?utm_source=chatgpt.com)	Теми 1-5.
CHI Software / CHI Academy Курс ML Engineer — У програмі є модуль «Навчання з підкріпленням та Q-навчання» як частина підготовки ML-інженера (https://chiacademy.software/courses/machine-learning?utm_source=chatgpt.com).	Тема 9
Coursera Fundamentals of Reinforcement Learning Вводный курс в специализации: изучение MDP, trade-off exploration vs exploitation, основные алгоритмы (https://www.coursera.org/learn/fundamentals-of-reinforcement-learning?utm_source=chatgpt.com).	Теми 1-6.
Coursera Deep Learning and Reinforcement Learning (IBM) - Курс, сочетающий глубинное обучение и RL: нейронные сети, обучение с подкреплением, практическая часть (https://www.coursera.org/learn/deep-learning-reinforcement-learning?utm_source=chatgpt.com).	Теми 6-10

Здобувач може самостійно обрати ресурси, які будуть відповідати навчальній дисципліні. Для зарахування в якості періодичного, поточного або фрагмента підсумкового контролю здобувач має подати документ, що підтверджує неформальну освіту, програму, зробити опис компетентностей та результатів навчання, сформованих під час неформальної освіти.

Викладач залишає за собою право перевірити сформованість перелічених компетентностей та результатів навчання

9. Методи навчання

Словесні: лекції (проблемні, лекції-візуалізації, лекції з аналізом конкретних ситуацій та ін.),

Наочні: ілюстрація (у тому числі мультимедійні презентації), демонстрація, презентація результатів власних досліджень.

Практичні: розв'язання розрахункових задач; практичні роботи; лабораторні роботи; виконання дослідів;

Самостійна робота студентів: опрацювання теоретичних джерел, розв'язання задач, підготовка міні-звітів і презентацій, виконання індивідуальних завдань з аналізу показників різних алгоритмів навчання.

10. Форми контролю і методи оцінювання

Контроль здійснюється з дотриманням вимог об'єктивності, індивідуального підходу, систематичності і системності та всебічності.

поточний контроль – опитування; оцінювання індивідуальної роботи, доповідей, PowerPoint презентацій; письмових завдань різних типів (вправ, анотацій, оформлення бібліографії та посилань);

підсумковий контроль – залік.

10.1. Критерії оцінювання на підсумковому контролі:

	100 бальна шкала / Оцінка ECTS	Теоретична підготовка	Практична підготовка
Зараховано	90-100/ А	Здобувач у повному обсязі володіє навчальним матеріалом, вільно, розгорнуто, обґрунтовано та аргументовано його викладає під час усних виступів та письмових відповідей. Здобувач демонструє чітке знання відповідних категорій, їх змісту, розуміння їх взаємозв'язку, правильно формулює тлумачення відповідних понять, демонструє знання змісту передбачених програмою робить самостійні висновки. Здобувач вміє самостійно знаходити додаткову інформацію та використовувати її для реалізації поставлених завдань, вільно використовує нові інформаційні технології для поповнення знань.	Здобувач може аргументовано обрати раціональний спосіб виконання практичних завдань, виконує практичні завдання не передбачені навчальною програмою, вільно використовує набуті теоретичні знання при аналізі практичного матеріалу, проявляє творчий підхід до виконання індивідуальних та колективних завдань при самостійній роботі.
Зараховано	85 – 89 / В	Здобувач достатньо повно володіє навчальним матеріалом, обґрунтовано його викладає під час усних виступів та письмових відповідей, використовуючи при цьому обов'язкову літературу, застосовує знання для розв'язання стандартних ситуацій, самостійно аналізує, узагальнює і систематизує навчальну інформацію, але допускає неточності, які не є суттєвими для характеристики предмету питання та не впливають істотно на загальну	Здобувач має стійкі навички виконання практичних завдань, правильно вирішує більшість практичних завдань.

		характеристику того чи іншого явища (поняття).	
	75 – 84 / C	Здобувач виявляє загалом високий рівень знань щодо всієї програми навчальної дисципліни, на достатньому рівні володіє навчальним матеріалом, застосовує знання для розв'язання стандартних ситуацій, але не вміє самостійно аналізувати деякі питання, не повністю переконливо аргументувати свої відповіді, допускає незначні неточності.	Здобувач за зразком самостійно виконує практичні завдання, передбачені програмою навчальної дисципліни.
Зараховано	70 – 74 / D	Здобувач володіє навчальним матеріалом на репродуктивному рівні або відтворює певну частину навчального матеріалу з елементами логічних зв'язків. Здобувач знає основні поняття навчального матеріалу, але має ускладнення під час виділення суттєвих ознак вивченого та під час виявлення причинно-наслідкових зв'язків та формулювання висновків.	Здобувач має елементарні навички виконання практичних завдань, правильно вирішує лише половину практичних завдань.
Зараховано	60 – 69 / E	Здобувач не повною мірою розуміє предмет навчальної дисципліни, наявні недоліки у розкритті змісту понять. Здобувач надає нечіткі характеристики відповідних явищ, викладає свої думки з істотним порушенням логіки подання матеріалу.	Здобувач може використовувати знання лише в стандартних практичних ситуаціях, має нестійкі навички виконання практичних завдань, робить багато суттєвих помилок.
Не зараховано з можливістю повторного складання/	35 – 59 / FX	Здобувач неправильно розкриває сутність базових питань навчальної дисципліни, допускає суттєві змістові помилки, володіє навчальним матеріалом поверхнево й фрагментарно, безсистемне виокремлює випадкові ознаки вивченого, не вміє сформулювати свою думку та викласти її в логічній послідовності, робити узагальнення та висновки.	Здобувач здатний виконати лише окремі практичні завдання за допомогою викладача. У здобувача відсутні сформовані уміння та навички.
Не зараховано з обов'язковим повторним вивченням дисципліни/	0 – 34 / F	Здобувач не знає основних положень навчальної дисципліни, не володіє навчальним матеріалом.	Здобувач виконує лише елементи практичних завдань, потребує постійної допомоги викладача.

11. Питання до підсумкового контролю

Тема 1.

1. У чому полягає основна відмінність навчання з підкріпленням від навчання з учителем та без учителя?
2. Які ключові компоненти системи навчання з підкріпленням?
3. Як поняття “агент” і “середовище” визначаються в контексті RL?
4. Чому RL вважається містком між машинним навчанням та теорією керування?
5. Які задачі штучного інтелекту найкраще розв'язуються за допомогою RL?

Тема 2.

6. Яке призначення середовища OpenAI Gym у навчанні з підкріпленням?
7. Які основні методи реалізовані в стандартному API Gym?

6. Що таке простір спостережень (observation space) і простір дій (action space) у Gym?
7. Як відбувається взаємодія агента із середовищем у Gym?
8. Які приклади середовищ із Gym використовуються для навчання базових RL-алгоритмів?

Тема 3.

11. Які елементи визначають Марковський процес прийняття рішень?
12. Що означає властивість марковськості?
13. Як виглядає формальне визначення MDP через множини S, A, P, R, γ , A, P, R, γ ?
14. Яке значення має коефіцієнт дисконтування γ у MDP?
15. Як у MDP визначається оптимальна стратегія (policy)?

Тема 4.

16. Яке місце займає динамічне програмування в теорії RL?
17. Як формулюється рівняння Беллмана для цінності стану?
18. Чим відрізняються рівняння для функцій цінності стану та дії?
19. Чому динамічне програмування вимагає повної моделі середовища?
20. Які обмеження має класичне динамічне програмування в задачах RL?

Тема 5.

21. У чому полягає принцип ітерації по цінностях?
22. Як відбувається процес оновлення функції цінності в ітерації по цінностях?
23. У чому полягає ітерація по стратегіях і які її етапи?
24. Як співвідносяться обидва методи — ітерація по цінностях і по стратегіях?
25. Які критерії зупинки використовуються в ітераційних методах?

Тема 6.

26. У чому полягає ідея методів Монте-Карло в RL?
27. Чому ці методи не потребують знання моделі середовища?
28. Як оцінюється функція цінності за методом Монте-Карло?
29. У чому полягає відмінність між епізодичним та безперервним Монте-Карло?
30. Які обмеження мають Монте-Карло методи в порівнянні з TD-підходами?

Тема 7.

31. У чому полягає дилема «дослідження проти використання»?
32. Які основні підходи застосовуються для балансування дослідження й експлуатації?
33. Як працює ϵ -жадібна стратегія (epsilon-greedy)?
34. Які недоліки має жадібний підхід до вибору дії?
35. Як можна оцінити ефективність політики з точки зору балансу exploration/exploitation?

Тема 8.

36. Яка основна ідея TD-навчання?
37. У чому полягає різниця між методами Монте-Карло та TD?
38. Як формується TD-помилка (temporal difference error)?
39. Як відбувається оновлення функції цінності у TD(0)?
40. Чому TD-методи вважаються більш ефективними для онлайн-навчання?

Тема 9.

41. Які п'ять етапів включає оновлення алгоритму SARSA?
42. Як SARSA відрізняється від Q-learning?
43. Чому SARSA вважається “on-policy” алгоритмом?
44. Як ϵ -жадібна стратегія інтегрується в SARSA?
45. У яких типах середовищ SARSA може бути ефективнішим за Q-learning?

Тема 10.

46. У чому полягає ідея поєднання Q-learning із нейронними мережами?
47. Яку роль відіграє буфер досвіду (experience replay) у DQN?

48. Для чого використовується “target network” у DQN?
 49. Які основні проблеми стабільності виникають у DQN-навчанні?
 50. Як відбувається оновлення ваг нейронної мережі під час DQN-процесу?

12. Розподіл балів, які отримують студенти

Поточне тестування та самостійна робота										Сума
Змістовий модуль №1					Змістовий модуль № 2					
T1	T2	T3	T4	T5	T1	T2	T3	T4	T5	100
10	10	10	10	10	10	10	10	10	10	

T1, T2 ... – теми змістових модулів.

Шкала оцінювання: національна та ECTS

Загальна сума балів	Оцінка ECTS	Національна шкала	
90 — 100	A – «відмінно»	5 «відмінно»	« з а л і к »
85 — 89	B – «дуже добре»	4 «добре»	
75 — 84	C – «добре»		
70 — 74	D – «задовільно»	3 «задовільно»	
60 — 69	E – «допустимо»		
35 — 59	F – «незадовільно з можливістю повторного складання»	2 «незадовільно»	« н е з а л і к »
0 — 34	FX – «незадовільно з обов’язковим повторним курсом»		

13. Навчально-методичне забезпечення

Презентаційні матеріали в електронному форматі.

14. Рекомендована література

14.1 Основна література

1. Sutton, R. S., Barto, A. G. *Reinforcement Learning: An Introduction (2nd Edition)*. The MIT Press 2018– 548 p. Режим доступу: <https://dokumen.pub/qdownload/reinforcement-learning-an-introduction-2nbsped-0262039249-9780262039246.html>
2. Andrea Lonza *Reinforcement Learning Algorithms with Python* Packt Publishing, 2019 – 356 p, Режим доступу: <https://dokumen.pub/hands-on-reinforcement-learning-for-games-1839214937-9781839214936.html>

14.2 Допоміжна література

3. François-Lavet, V., Henderson, P., Islam, R. et al.. *An Introduction to Deep Reinforcement Learning*. 2018 Режим доступу: <https://arxiv.org/abs/1811.12560>
4. Busoniu, L., Babuska, R., De Schutter, B., & Ernst, D. *Reinforcement Learning and dynamic programming using function approximators* 2010, 282 p. - Режим доступу: <https://orbi.uliege.be/bitstream/2268/27963/1/book-FA-RL-DP.pdf>.
5. Laura Graesser L., Keng W. *Foundations of Deep Reinforcement Learning: Theory and Practice in Python*, Addison-Wesley Professional, 2019, 416 p. - Режим доступу: <https://dokumen.pub/foundations-of-deep-reinforcement-learning-theory-and-practice-in-python-1nbsped-9780135172384-0135172381.html>
6. Szepesvári, C. *Algorithms for Reinforcement Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning*
DOI:10.2200/S00268ED1V01Y201005AIM009 Morgan & Claypool Publishers. 2010
- Режим доступу:
https://www.researchgate.net/publication/220696313_Algorithms_for_Reinforcement_Learning

14.3 Електронні інформаційні ресурси

7. OpenAI Spinning Up in Deep RL (2018). <https://spinningup.openai.com/en/latest/>
8. Gymnasium: An API standard for reinforcement learning with a diverse collection of reference environments <https://gymnasium.farama.org/index.html>