



Національний технічний університет України
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»



Кафедра обчислювальної
техніки

Навчання з підкріпленням

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	Другий (магістерський)
Галузь знань	F Інформаційні технології
Спеціальність	F2 Інженерія програмного забезпечення, F7 Комп'ютерна інженерія, F6 Інформаційні системи та технології
Освітня програма	Інженерія програмного забезпечення комп'ютерних систем, Комп'ютерні системи та мережі
Статус дисципліни	Вибіркова
Форма навчання	очна(денна)
Рік підготовки, семестр	2 курс, осінній семестр
Обсяг дисципліни	4 кредити/120 год. денна форма: лекцій 36 годин, Пр. роб. 18 год., СРС 66 год.
Семестровий контроль/ контрольні заходи	Залік
Розклад занять	Згідно розкладу на осінній семестр поточного навчального року за адресою http://rozklad.kpi.ua
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектори: ст. викладач Кочура Юрій Петрович, iuriy.kochura@gmail.com Практичні роботи: ст. викладач Кочура Юрій Петрович, iuriy.kochura@gmail.com
Розміщення курсу	https://courses-cs-kpi.github.io/

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Навчання з підкріпленням (англ. reinforcement learning, RL) — це тип машинного навчання для автоматизованого прийняття рішень на основі взаємодій: агент навчається приймати оптимальні рішення, беручи до уваги винагороди, які отримує від середовища у ході взаємодії. Починаючи з 2014 року глибоке навчання з підкріпленням (deep RL) зазнало значних успіхів та стало одним з найінтенсивніших напрямків досліджень у галузі штучного інтелекту. Сьогодні навчання нейронних мереж з підкріпленням активно використовується у робототехніці, фінансовій сфері та іграх. У ряді завдань досягнуто надлюдської продуктивності: відео ігри, покер, а також у настільних іграх, включаючи го та шахи.

Цей курс познайомить Вас з сімейством статистичних алгоритмів, які вивчають оптимальну стратегію, метою якої є максимізація загальної винагороди, отриманої агентом при взаємодії з навколишнім середовищем.

Предмет навчальної дисципліни – методи та засоби навчання з підкріпленням, які спрямовані на моделювання процесів прийняття рішень агентами з метою максимізації сумарної винагороди у різних середовищах.

Навчальна мета

- Познайомитись з фундаментальною теорією та особливостями RL;
- Навчитись вирішувати прості задачі за допомогою методів RL;
- Познайомитись з дослідженнями в галузі RL ($\leq$ останні 10 років).

Вивчення дисципліни підсилює наступні загальні та фахові компетенції:

- Здатність вчитися і оволодівати сучасними знаннями
- Здатність використовувати сучасні методи і мови програмування для розроблення алгоритмічного та програмного забезпечення
- Здатність розробляти, адаптувати, використовувати програмно-технічні засоби для збирання, оброблення та інтелектуального аналізу даних в комп'ютерних системах: паралельних, розподілених, хмарних, безпечних, інтелектуальних, розумних, Інтернет речей, зокрема, з використанням штучного інтелекту.

У відповідності до вищезазначеного, підсилені загальні та фахові компетенції дадуть наступні програмні результати навчання:

1. Вміння застосовувати знання для ідентифікації, формулювання і розв'язування технічних задач спеціальності, використовуючи методи, що є найбільш придатними для досягнення поставлених цілей.
2. Вміння системно мислити та застосовувати творчі здібності до формування нових ідей.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Для опанування предмету «Навчання нейронних мереж з підкріпленням» потрібно володіти компетентностями з наступних дисциплін: «Вища математика», «Аналітична геометрія та лінійна алгебра», «Дискретна математика», «Програмування», «Системи штучного інтелекту», «Нейронні мережі».

Розглянута теорія та отриманий практичний досвід в рамках даної навчальної дисципліни буде корисним для розв'язання прикладних задач у робототехніці, автономному керуванні, фінансовому прогнозуванні, оптимізації бізнес-процесів та створенні інтелектуальних ігрових агентів.

3. Зміст навчальної дисципліни

Розділ 1. Вступ до навчання з підкріпленням.

Базова концепція «агент – середовище – винагорода». Основні підходи до навчання з підкріпленням. Приклади практичних застосувань.

Розділ 2. Марковські процеси.

Марковські процеси станів, процеси винагороди та процеси прийняття рішень. Властивості марковських процесів. Математичний апарат для опису взаємодії агента із середовищем.

Розділ 3. Планування за допомогою динамічного програмування.

Методи обчислення оптимальних стратегій. Принцип Беллмана. Ітерація цінності та ітерація стратегії. Обчислювальна складність та практичні обмеження.

Розділ 4. Методи навчання Монте-Карло та методи з часовими різницями (TD-learning).

Основи стохастичного моделювання. Оцінка функції цінності методом Монте-Карло. Підходи TD(0), TD(λ). Порівняння та поєднання методів.

Розділ 5. Методи апроксимації функції цінності.

Лінійні та нелінійні методи апроксимації. Використання функційних наближень для роботи у великих або безперервних просторах станів. Використання нейронних мереж.

Розділ 6. Градієнтні методи стратегій (policy gradient methods).

Пряме оптимізаційне навчання стратегій. Стохастичні градієнтні методи. Переваги та обмеження підходів на основі стратегії.

4. Навчальні ресурси та матеріали

4.1. Базова література

1. Кочура, Ю. П. Нейронні мережі [Електронний ресурс] : підручник для здобувачів ступеня бакалавра/магістра за спец. 123 «Комп'ютерна інженерія» та 121 «Інженерія програмного забезпечення» / Ю. П. Кочура, Ю. Г. Гордієнко ; КПІ ім. Ігоря Сікорського. – Електрон. текст. дані (1 файл: 3,98 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2024. – 114 с. – (Серія «Штучний інтелект для індустрії»). – Назва з екрана. URL <https://ela.kpi.ua/handle/123456789/73660>

4.2. Додаткова література

2. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
3. Zai, A., & Brown, B. (2020). *Deep reinforcement learning in action*. Manning Publications.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компоненту)

5.1. Лекційні заняття

№ лекції	Назва лекції та перелік основних питань	Кількість ауд. годин
1	Вступ до навчання з підкріпленням. Базова концепція «агент – середовище – винагорода». Основні підходи до навчання з підкріпленням. Приклади практичних застосувань.	4
2	Марковські процеси. Марковські процеси станів, процеси винагорода та процеси прийняття рішень. Властивості марковських процесів. Математичний апарат для опису взаємодії агента із середовищем.	4

3	Планування за допомогою динамічного програмування. Методи обчислення оптимальних стратегій. Принцип Беллмана. Ітерація цінності та ітерація стратегії. Обчислювальна складність та практичні обмеження.	8
4	Методи навчання Монте-Карло та методи з часовими різницями (TD-learning). Основи стохастичного моделювання. Оцінка функції цінності методом Монте-Карло. Підходи TD(0), TD(λ). Порівняння та поєднання методів.	8
5	Методи апроксимації функції цінності. Лінійні та нелінійні методи апроксимації. Використання функційних наближень для роботи у великих або безперервних просторах станів. Використання нейронних мереж.	8
6	Градiєнтні методи стратегій (policy gradient methods). Пряме оптимізаційне навчання стратегій. Стохастичні градієнтні методи. Переваги та обмеження підходів на основі стратегії.	4
	Разом:	36

5.2. Лабораторні/практичні заняття

№ з/п	Назва лабораторної/практичної роботи	Кількість ауд. годин
1	Навчання агента у середовищах різної складності	6
2	Компроміс між отриманням нових знань та використанням раніше набутого досвіду	6
3	Передбачення найкращих станів та дій агента	6
	Разом:	18

6. Самостійна робота здобувача вищої освіти денної форми навчання

До СРС відносяться такі види робіт:

- вивчення теоретичного матеріалу лекційних занять

- виконання практичних робіт та оформлення протоколів
- підготовка матеріалів на семінарські заняття
- підготовка до контрольної роботи та семестрового контролю

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Академічна доброчесність

Під час проходження цього курсу студенти зобов'язані дотримуватись політики та принципів академічної доброчесності визначених Кодексом честі КПІ ім. Ігоря Сікорського (Детальніше Розділ 3: <https://kpi.ua/code>) та усі наступні правила:

1. Кожен з Вас повинен відправляти на перевірку власно виконану роботу. Використання чужих розв'язків або програмного коду і представлення їх за свої напрацювання є плагіатом та серйозним порушенням основних академічних стандартів.
2. Ви не повинні ділитися своїми розв'язками з іншими студентами, а також просити інших ділитися своїми розв'язками з Вами.
3. Якщо Ви отримували допомогу у вирішенні певного завдання, Ви маєте зазначити це у звіті, а саме: від кого та яку допомогу отримали.

Норми етичної поведінки

Норми етичної поведінки студентів і працівників університету визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Виконання практичних завдань

Викладачем встановлюється граничний термін для виконання кожного практичного завдання з метою послідовного та збалансованого засвоєння студентами навчального матеріалу, визначеного даним силабусом. Здача на перевірку практичних завдань після встановлених термінів супроводжується штрафними балами. Кінцевим терміном для здачі на перевірку усіх практичних завдань є останнє заняття з комп'ютерного практикуму. У період проведення в університеті семестрового контролю приймання практичних завдань не здійснюється. Студент(-ка) має можливість переробити або доопрацювати практичне завдання на перездачі з метою підвищення свого поточного рейтингу.

Оцінювання контрольних заходів у разі пропуску

У разі пропуску студентом(-кою) модульної контрольної роботи, студент(-ка) отримує нуль балів за цей контрольний захід. У такому випадку, студент(-ка) має можливість написати пропущену модульну контрольну роботу при наявності відповідного документа, який засвідчує поважну причину відсутності на цьому заході. Повторне проходження модульної контрольної роботи для покращення свого результату непередбачено.

Процедура оскарження оцінок

Якщо студент(-ка) вважає, що його(її) роботу недооцінено або переоцінено — потрібно звернутися до викладача, який здійснював оцінювання та повідомити про це із зазначенням короткого обґрунтування (оцінка буде переглянута).

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Календарний контроль: проводиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу

Семестровий контроль: Залік

Система оцінювання

30 балів	Практичні завдання (10 балів кожне)
20 балів	Семінар
10 балів	Контрольна
40 балів	Залік

Важливо! Умова допуску до семестрового контролю (заліку):
Практичні завдання + Семінар + Контрольна $\geq$ 36 балів

Шкала оцінок КПІ ім. Ігоря Сікорського

A = 95–100	Відмінно
B = 85–94	Дуже добре
C = 75–84	Добре
D = 65–74	Задовільно
E = 60–64	Достатньо
F < 60	Незадовільно
Fx < 36	Не допущений
Порушення Кодексу честі	Усунений

9. Нормативні положення політики навчальної дисципліни (освітнього компонента)

Правила відвідування занять. Відвідування занять є вільним, бали за присутність на лекціях та практичних заняттях не нараховуються, відповідно штрафні бали у разі відсутності на занятті не передбачаються.

Дистанційний режим навчання. У разі запровадження обмежень на відвідування університету, пов'язаних з введенням карантину або режиму воєнного стану в державі, освітній процес здійснюється у дистанційному режимі відповідно до Положення про дистанційне навчання в КПІ ім. Ігоря Сікорського (<https://osvita.kpi.ua/index.php/node/188>), Регламенту організації освітнього процесу в дистанційному режимі (<https://profkom.kpi.ua/reglament-organizatsiyi-osvitnogo-protsesu-v-distantsynomu-rezhimi>) та Регламенту проведення семестрового контролю в дистанційному режимі (<https://osvita.kpi.ua/node/148>). У режимі дистанційного навчання заняття відбуваються у вигляді онлайн-конференції на платформах BigBlueButton, Google Meet, Zoom. Посилання на конференцію видається на початку семестру і розміщується в АС «Електронний кампус». З метою забезпечення якісної підготовки здобувачів, дистанційний курс дисципліни розміщено на Платформа дистанційного навчання «Сікорський» (<https://www.sikorsky-distance.org>). Результати оцінювання висвітлюють у АС «Електронний кампус» на особистій сторінці здобувача (<https://ecampus.kpi.ua>).

Правила поведінки на заняттях. На заняттях слід дотримуватись норм етичної поведінки визначених у Кодексі честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (<https://kpi.ua/code>). На території університету здобувачі мають поводити себе відповідно до Правил внутрішнього розпорядку (<https://kpi.ua/admin-rule>).

Визнання результатів навчання, набутих у неформальній/інформальній освіті. Положення про визнання результатів навчання, набутих у неформальній / інформальній освіті (<https://osvita.kpi.ua/index.php/node/179>) регламентує визнання результатів навчання, набутих у неформальній/інформальній освіті:

- у разі навчання на неформальній освіті, яку студент обрав самостійно і хоче зарахувати результати навчання в рамках дисципліни, проходиться процедура валідації, що передбачає подання здобувачем заяви на ім'я декана, декларації підтверджувальних документів. Рішення про визнання чи не визнання приймається комісією у складі завідувача кафедри, викладача, гаранта освітньо-професійної програми.

Політика використання штучного інтелекту. Використання штучного інтелекту регламентується «Політикою використання штучного інтелекту для академічної діяльності в КПІ ім. Ігоря Сікорського» (<https://osvita.kpi.ua/node/1225>). Усі завдання, як під час виконання навчальних завдань з дисципліни, так і індивідуальні завдання, мають бути результатом власної оригінальної роботи здобувача. Використання ШІ для автоматичної генерації відповідей без подальшого їх аналізу та доопрацювання заборонено.

Робочу програму навчальної дисципліни (силабус)

Складено, ст. викладач кафедри обчислювальної техніки Кочура Ю. П.

Ухвалено кафедрою обчислювальної техніки (протокол № 12 від 23.06.2025р.).

Погоджено методичною комісією ФІОТ (протокол № 11 від 27.06.2025 р.).