№ 4(58) 2026
НАУКА і ТЕХНІКА СЬОГОДНІ
ISSN 2786-6025 Online
УДК 004.94:004.8:007.52
https://doi.org/10.52058/2786-6025-2026-4(58)-2765-2781

Белзецький Руслан Станіславович кандидат технічних наук, доцент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, https://orcid.org/0000-0003-1574-8831

Добровольська Євгенія Романівна студентка групи 1КН-22б кафедри комп'ютерних наук, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, https://orcid.org/0009-0008-5900-4731

Чернокнижник Артем Віталійович студент групи 1КН-24б кафедри комп'ютерних наук, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, https://orcid.org/0009-0001-9532-3624

РОЗРОБКА ПРОГРАМНО-АПАРАТНОГО КОМПЛЕКСУ КЕРУВАННЯ МОБІЛЬНИМ РОБОТОМ НА ОСНОВІ АЛГОРИТМУ НАВЧАННЯ З ПІДКРІПЛЕННЯМ В СЕРЕДОВИЩІ WEBOTS

Анотація. У статті розглянуто застосування методів навчання з підкріпленням для розв’язання задачі автономної навігації мобільних роботів у складному динамічному середовищі. Зростання складності робототехнічних систем та потреба у високому рівні автономності зумовлюють необхідність використання інтелектуальних алгоритмів керування, здатних адаптуватися до невизначених умов. Особливу увагу приділено методам глибокого навчання з підкріпленням (DRL), які поєднують можливості нейронних мереж і підходів навчання через взаємодію з середовищем. У роботі проведено аналіз сучасних наукових досліджень і публікацій, присвячених використанню DRL для задач планування траєкторії, уникнення перешкод і побудови ефективної навігаційної стратегії мобільних роботів.

Запропоновано підхід до навчання робота навігації у віртуальному середовищі моделювання Webots, що дозволяє безпечно та ефективно проводити експериментальні дослідження. Для реалізації моделі використано алгоритми навчання з підкріпленням, що дає змогу поступово формувати оптимальну стратегію поведінки на основі критерію оптимальності. Навчання відбувається шляхом багаторазової взаємодії робота із середовищем, у процесі якої система накопичує досвід і покращує якість прийняття рішень.

ISSN 2786-6025 Online

У роботі описано структуру моделі, принцип формування функції підкріплення, а також етапи навчання та тестування системи. Проведено серію експериментів у симуляційному середовищі Webots з метою оцінювання ефективності запропонованого підходу. Аналіз результатів експериментів показує стабілізацію поведінки агента в процесі навчання. Отримані результати підтверджують здатність моделі ефективно знаходити шлях до цільової точки, уникаючи перешкод і оптимізуючи траєкторію руху.

Запропонований підхід демонструє перспективність використання методів глибокого навчання з підкріпленням для задач автономної навігації мобільних роботів. Результати дослідження можуть бути використані для подальшого вдосконалення інтелектуальних систем керування роботами, а також для розробки більш складних робототехнічних систем, здатних працювати в реальних умовах.

За результатами, представленими у роботі, отримано свідоцтво про державну реєстрацію авторського права на комп'ютерну програму «Контролер керування мобільним роботом у середовищі Webots на основі алгоритму навчання з підкріпленням» № 145166 від 3.04.2026 [1].

Ключові слова: навчання з підкріпленням, Webots, автономна навігація, мобільна робототехніка, фізичне моделювання.

Belzetskyi Ruslan Stanislavovych Candidate of Technical Sciences, Associate Professor of the Department of Computer Science, Vinnytsia National Technical University, Vinnytsia, https://orcid.org/0000-0003-1574-8831

Dobrovolska Yevheniya Romanivna Student of Group 1КН-22б of the Department of Computer Science, Faculty of Intellectual Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, https://orcid.org/0009-0008-5900-4731

Chernoknyzhnyk Artem Vitaliyovych Student of Group 1КН-24б of the Department of Computer Science, Faculty of Intellectual Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, https://orcid.org/0009-0001-9532-3624

DEVELOPMENT OF A SOFTWARE AND HARDWARE CONTROL SYSTEM FOR A MOBILE ROBOT BASED ON A REINFORCEMENT LEARNING ALGORITHM IN THE WEBOTS ENVIRONMENT

Abstract. The article considers the application of reinforcement learning methods for solving the problem of autonomous navigation of mobile robots in a

ISSN 2786-6025 Online

complex dynamic environment. The increasing complexity of robotic systems and the need for a high level of autonomy necessitate the use of intelligent control algorithms capable of adapting to uncertain conditions. Special attention is paid to deep reinforcement learning (DRL) methods, which combine the capabilities of neural networks and learning approaches through interaction with the environment. The paper analyzes current research and publications devoted to the use of DRL for trajectory planning, obstacle avoidance, and developing effective navigation strategies for mobile robots.

An approach to training a robot for navigation in the Webots simulation environment is proposed, which allows safe and efficient experimental research. Reinforcement learning algorithms are used to implement the model, enabling the robot to gradually form an optimal behavior strategy based on an optimality criterion. Training occurs through repeated interaction of the robot with the environment, during which the system accumulates experience and improves decision-making quality.

The paper describes the model structure, the principle of forming the reinforcement function, as well as the stages of training and testing the system. A series of experiments was conducted in the Webots simulation environment to evaluate the effectiveness of the proposed approach. Analysis of the experimental results shows stabilization of the agent's behavior during training. The obtained results confirm the model's ability to effectively find a path to the target point while avoiding obstacles and optimizing the trajectory.

The proposed approach demonstrates the promise of using deep reinforcement learning methods for autonomous navigation tasks of mobile robots. The research results can be used for further improvement of intelligent robot control systems, as well as for the development of more complex robotic systems capable of operating in real-world conditions.

Based on the results presented in the paper, a certificate of state registration of copyright for the computer program "Controller for controlling a mobile robot in the Webots environment based on a reinforcement learning algorithm" No. 142304 dated April 10, 2026 was obtained [1].

Keywords: reinforcement learning, Webots, autonomous navigation, mobile robotics, physical simulation.

Постановка проблеми. Сучасний розвиток штучного інтелекту та машинного навчання зумовлює зростаючий інтерес до методів, здатних забезпечувати автономне прийняття рішень у складних та динамічних середовищах. Одним із таких підходів є навчання з підкріпленням (RL), у межах якого агент навчається обирати оптимальні дії шляхом взаємодії із середовищем та отримання сигналів критерію оптимальності. Поєднання цього

ISSN 2786-6025 Online

підходу з глибокими нейронними мережами сформувало напрям глибокого навчання з підкріпленням (DRL), що дозволяє розв'язувати багатовимірні задачі прийняття рішень.

Разом з тим застосування таких методів пов'язане з рядом наукових і практичних проблем, зокрема складністю процесу навчання, необхідністю значної кількості взаємодій із середовищем та забезпеченням стабільності роботи алгоритмів. Ці питання є особливо актуальними для таких галузей, як робототехніка, комп'ютерний зір, автономні транспортні системи та інтелектуальні системи керування.

Таким чином, актуальним науковим завданням є дослідження та вдосконалення методів навчання з підкріпленням для ефективного розв'язання задач прийняття рішень у складних середовищах та їх практичного застосування в сучасних системах штучного інтелекту.

Аналіз досліджень і публікацій. У сучасних наукових дослідженнях використання методів навчання з підкріпленням (RL) в контексті вирішення завдань навігації автономних мобільних роботів є актуальним та перспективним. Останніми роками спостерігається зростаюча тенденція до застосування глибокого навчання з підкріпленням (DRL) для планування шляху мобільних роботів у приміщеннях, що супроводжується значними успіхами [2].

Глибоке навчання з підкріпленням (DRL) вперше було застосовано для дискретного керування мобільними роботами з метою реалізації функції уникнення перешкод. У дослідженні Пфайффера та ін. [3] для навчання планувальника шляху на основі згорткових нейронних мереж (CNN) було використано результати роботи класичних алгоритмів Дейкстри та DWA як навчальні мітки. Такий підхід дозволив поєднати точність геометричних методів із гнучкістю нейромережевих архітектур.

Паралельно з цим, у працях Л. Тай та інші [4] запропоновано наскрізну (end-to-end) стратегію уникнення перешкод на основі глибокого Q-навчання (DQN), а також представлено систему навігації, що базується на розріджених сигналах лазерного радара без використання мап середовища.

Окрему групу підходів становлять гібридні методи, спрямовані на реалізацію навігації на великі відстані. Зокрема, у дослідженнях Фауст та інші [5] успіху було досягнуто шляхом інтеграції імовірнісних дорожніх карт (PRM) із алгоритмами RL. Аналогічний підхід розвивають Френсіс та інші [6], де PRM виступає як планувальник на основі вибірки, а система AutoRL забезпечує адаптивність керування в складних середовищах.

Значна увага в публікаціях приділяється вибору інструментарію моделювання. Як зазначається в роботі Мішель, яка присвячена платформі Webots [7], зростаюча складність завдань RL зумовлює потребу в точних 3D-симуляторах із відкритим кодом. Webots надає середовища, які можна повністю

ISSN 2786-6025 Online

налаштовувати, а також забезпечує високоточне моделювання з реалістичною графікою та повну сумісність з операційною системою для роботів (ROS).

Результати сучасних досліджень підтверджують, що платформа Webots належить до найнадійніших і ефективних засобів для проєктування й апробації алгоритмів глибокого навчання з підкріпленням. Висока достовірність одержуваних результатів досягається завдяки інтеграції з реалістичним фізичним рушієм (ODE), що забезпечує коректне відтворення контактних взаємодій і складної поведінки мобільних роботів.

Загалом аналіз досліджень свідчить про відсутність універсального методу, який би забезпечував максимальну збіжність алгоритмів у надскладних середовищах. Це зумовлює актуальність подальших розробок, спрямованих на модернізацію існуючих методів DRL та створення нових архітектурних рішень для підвищення автономності мобільних роботів.

Мета статті – розроблення та дослідження програмно-апаратного комплексу керування мобільним роботом на основі алгоритму навчання з підкріпленням у середовищі Webots з функцією адаптивного огинання перешкод.

Виклад основного матеріалу. Ядро Webots базується на надійному та потужному фізичному двигуні (ODE). Його основними компонентами є двигун моделювання динаміки твердого тіла та двигун виявлення зіткнень. Розробка моделі включає кілька фаз: починаючи від моделювання деталей, визначення обмежувальних об'єктів, налаштування фізичних властивостей.

Webots представляє сцени з деревоподібною структурою, в якій кореневим вузлом є світ, а його дочірніми вузлами є елементи у світі. Отже, вузол робота повинен бути під кореневим вузлом і має містити контролер. Контролери – це скрипти, що відповідають за функціональність вузла. Для цілей RL необхідно включити спеціальний вузол-супервізор, який має повне знання про світ і може отримувати інформацію та змінювати будь-який інший вузол. Вузол RotationalMotor використовується для створення обертального руху навколо вибраної осі.

Для детекції перешкод у розробленій моделі мобільного робота використано вузли DistanceSensor. Моделювання роботи цих пристроїв реалізовано шляхом виявлення зіткнень вимірювальних променів датчиків із об'єктами віртуального середовища. Вузол GPS використовується для моделювання датчика глобального позиціонування, який отримує інформацію про своє абсолютне положення з програми контролера.

Прямий експорт CAD-моделей у середовище симуляції є неефективним через надлишкову деталізацію та невідповідність масштабів. Тому було використано Blender для масштабування та оптимізації геометрії. Також виконано коректне вирівнювання центрів мас коліс для забезпечення стабільної симуляції без артефактів. Вихідна модель містила складні елементи (різьба на

ISSN 2786-6025 Online

гвинтах, внутрішні порожнини моторів) (рис. 1), які не впливають на зовнішній вигляд, але перевантажують графічний конвеєр. Для зменшення цього впливу було застосовано методи ретопології та модифікатор Decimate, що дозволило зменшити кількість полігонів без втрати візуальної якості. Для коректного обертання коліс їхні центри мас були вирівняні суворо по геометричному центру осей обертання. Це критично важливо, оскільки зміщення PivotPoint у Webots призводить до «биття» колеса та нестабільного руху робота [8].

а) б)

Рисунок 1 – Розроблюваний робот: а – 3D модель; б – прототип

Для створення реалістичної програмно-апаратної моделі робота у Webots використовується механізм імпорту полігональних сіток, що дозволяє перенести складну геометрію з систем автоматизованого проєктування безпосередньо у віртуальний простір.

Процес імпорту у Webots реалізується через ієрархію вузлів Robot. На першому етапі до переліку дочірніх елементів зазначених вузлів додається компонент візуалізації Shape. У межах даного компонента геометричні характеристики об’єкта задаються шляхом ініціалізації поля geometry вузлом типу Mesh. Локалізація вихідних даних для побудови сітки визначається у полі URL вузла Mesh, де зазначається шлях до цільового файлу у форматі STL. Вузол Mesh використовується для завантаження оптимізованого файлу .obj з Blender, зберігши всі дрібні деталі: текстуру плати керування, ребра жорсткості шасі та специфічну форму дисків коліс.

Визначення візуальних характеристик поверхні моделі реалізується через поле appearance вузла Shape. Для задання фізично коректних параметрів матеріалу, таких як колір, текстура картування, коефіцієнти відбиття, використовуються спеціалізовані вузли. Зокрема, для реалістичного моделювання поверхонь на основі фізичних властивостей застосовується вузол PBRAppearance. Він визначає фізично обґрунтований візуальний вигляд вузла [9, 10].

Оскільки змодельовані форми є складними полігональними сітками, вони не можуть бути використані механізмами виявлення зіткнень. Це призведе до

ISSN 2786-6025 Online

занадто складних обчислень і уповільнить моделювання. Тому для розрахунку фізичних зіткнень у поле bounding Object було розміщено спрощені примітиви (Box, Cylinder) рисунок 2.

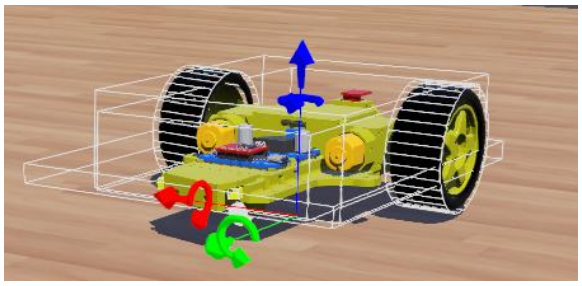


Рисунок 2 – Модель робота з його обмежувальними об'єктами

Базову структуру навчання з підкріпленням (RL) можна представити Марковським процесом прийняття рішень (MDP), який складається зі стану (S), дії (A), критерію оптимальності (R) та ймовірності переходу (P). На кожному кроці часу агент вибирає дію $a_t \in A$ у стані $s_t \in A$ і визначає розподіл ймовірностей дії $\pi(a|s)$. Тоді середовище забезпечує критерій оптимальності r_{t+1} і переходить до наступного стану s_{t+1}. Метою навчання з підкріпленням є вивчення оптимальної політики π^*, Шляхом максимізації кумулятивного критерію оптимальності:

$$R = \sum_{t=0}^{\infty} \gamma^t r_{t+1}$$

де γ – це коефіцієнт дисконтування, який контролює вплив майбутніх критеріїв оптимальності на поточні рішення. Цей базовий фреймворк широко застосовується для навчання оптимальних стратегій у динамічних середовищах [11].

У DRL різні алгоритми підходять для різних типів завдань, зокрема:

Q-навчання – це алгоритм на основі цінностей, який керує вибором дій шляхом вивчення функції цінностей $Q(s,a)$ для пар дій станів. Формула оновлення Q-навчання така:

$$Q(s_t,a_t) = Q(s_t,a_t) + \alpha(r_{t+1} + \gamma \operatorname*{Max}_{a'} Q\,(s_{t+1},a') - Q(s_t,a_t))$$

де α – швидкість навчання, а γ – коефіцієнт дисконтування. У DQN Q навчання поєднується з CNN для обробки високо вимірних вхідних зображень, що дозволяє агентам вивчати ефективні стратегії для складних візуальних завдань. Подвійне DQN зменшує переоцінку, відокремлюючи вибір дії від оцінки значення, роблячи модель стабільнішою [12].

Методи градієнта політики безпосередньо оптимізують політику $\pi(a|s,\theta)$, де θ – позначає параметри політики. Мета полягає в максимізації

ISSN 2786-6025 Online

очікуваного кумулятивного критерію оптимальності $J(\theta) = E_\pi[R]$, з формулою оновлення градієнта політики наступною:

$$\nabla_\theta J(\theta) = E_\pi[\nabla_\theta \, Log \, \pi(a|s,\theta) Q_\pi(s,a)]$$

Методи градієнта політики підходять для задач простору безперервної дії та забезпечують вищу ефективність обробки високовимірних безперервних просторів шляхом безпосередньої оптимізації політики [13].

Функція якості керування

Основою ефективного навчання алгоритму є розроблена функція винагороди, що відображає стан середовища у множину числових оцінок. Її значення залежить від:

1. досягнення цільових координат;
2. наявності зіткнень з перешкодами;
3. поточної відстані до цілі;
4. обраної траєкторії (рух по прямій чи ламаній);
5. витраченого часу.

Головною відмінністю розробленої функції є динамічний розрахунок оцінки наближення. На початкових етапах навчання використання виключно термінальних станів (досягнення цілі або зіткнення) є неефективним через низьку ймовірність випадкового виконання задачі.

Для оптимізації процесу збіжності алгоритму застосовано математичну модель експоненційного заохочення. Зі зменшенням відстані до цілі значення винагороди експоненційно зростає. Одночасно, для запобігання потраплянню алгоритму в локальні оптимуми, введено додаткові штрафи за зміну траєкторії та тривалість виконання. Це забезпечує поетапне та цілеспрямоване формування необхідної поведінки об'єкта керування.

Фрагмент коду обчислення функції винагороди

```
fitness = 0; // Поточне значення винагороди
norm_dist = ...; // Нормалізована відстань за даними GPS
action = ...; // Обрана дія (0-вперед,1/2 повороти)
// Перевірка термінальних станів
ifвідстань<порогу_досягненняthen
fitness = 25.0; // Винагорода за досягнення цілі
 return fitness, done = true;
if спрацював_сенсор_дотику then
 fitness = -5.0; // Штраф за зіткнення
 return fitness, done = true;
//Обчислення винагороди за наближення
ifnorm_dist< 42 then
// Визначення коефіцієнтів A(амплітуда) та g(крутизна)
fitness = A * (1 - exp(-g * (1 / norm_dist)));
else
 fitness = -norm_dist / 100.0; // Лінійний штраф за віддалення
```

ISSN 2786-6025 Online

```
// Застосування динамічних штрафів
if action != РУХ_ВПЕРЕД then
 fitness = fitness - 0.5; // Штраф за зміну траєкторії
fitness = fitness - 0.1; // Штраф за крок часу
return fitness;
```

Алгоритм навчання з підкріпленням, що реалізований в даній робот представлена на рисунку 3.

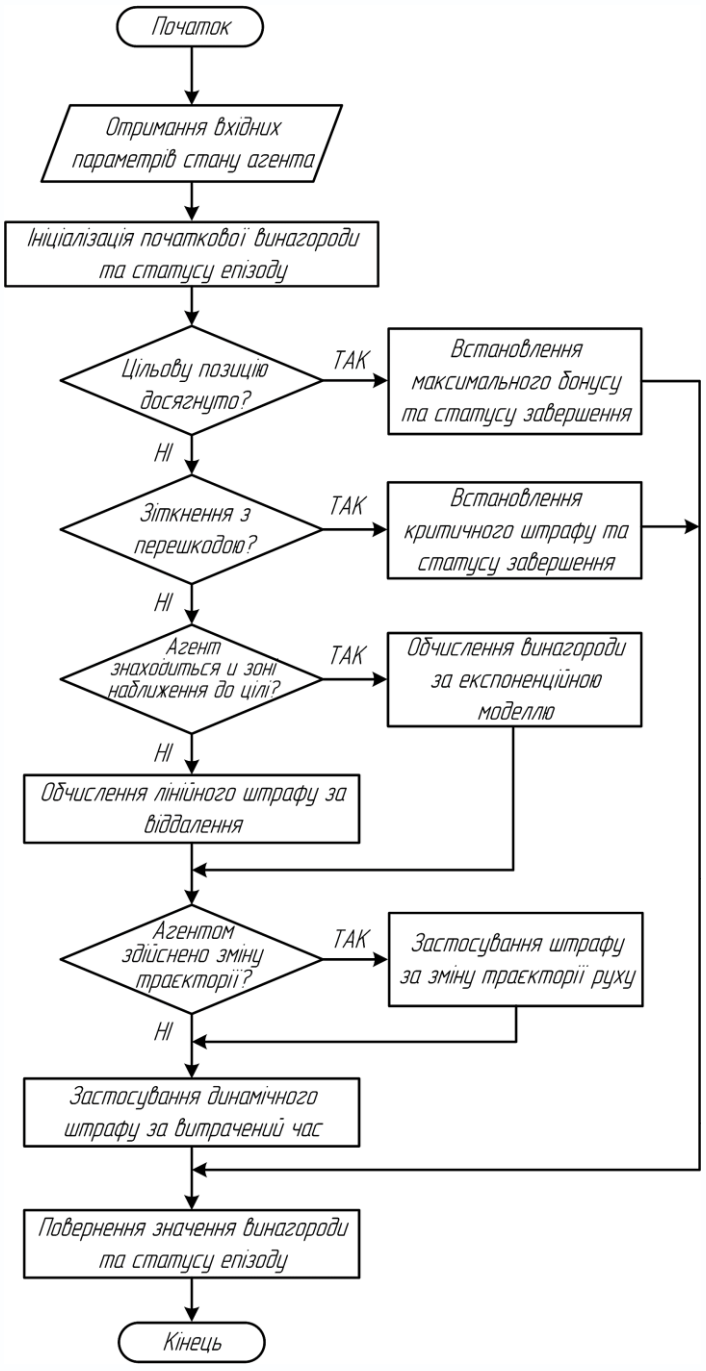


Рисунок 3 – Схема алгоритму навчання з підкріпленням

ISSN 2786-6025 Online

Інтеграція з середовищем моделювання

У середовищі Webots було розроблено віртуальний тренувальний полігон, призначений для дослідження алгоритмів автономної навігації мобільного робота. Структура полігону включає статичну перешкоду та визначену цільову зону, досягнення якої є основною задачею робота (рис. 4).

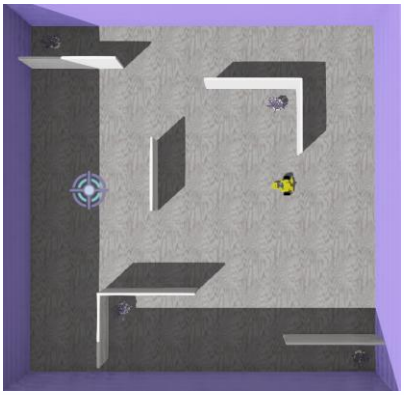


Рисунок 4 – Тренувальний полігон у середовищі Webots

З метою підвищення ефективності експериментальних досліджень було використано режим віртуального часу, який реалізований у Webots, що дозволяє запускати симуляції набагато швидше [7]. Такий підхід дозволяє значно скоротити тривалість навчання, забезпечуючи можливість проведення великої кількості ітерацій взаємодії з середовищем за певний час.

Навчання агента тривало 2000 епізодів. На початкових етапах поведінка мала стохастичний характер через активну фазу дослідження простору дій . В процесі оновлення параметрів нейронної мережі кількість випадкових дій зменшувалася, що супроводжувалося зростанням і подальшою стабілізацією функції підкріплення (рис. 5). Це свідчить про збіжність алгоритму та формування сталої політики керування.

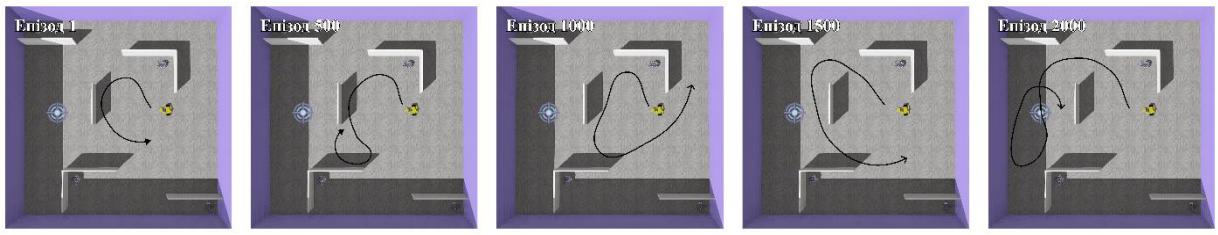


Рисунок 5 – Процес навчання робота

Для аналізу результатів навчання був розроблений скрипт analyze_ training_results.py, призначений для автоматизованого аналізу та візуалізації результатів навчання агента в середовищі Webots. Скрипт реалізовано мовою

ISSN 2786-6025 Online

Python, з використанням бібліотек *NumPy* для числових обчислень, pandas для роботи з табличними даними, matplotlib для побудови графіків. Архітектура скрипту побудована на основі класу TrainingAnalyzer, який інкапсулює всі необхідні методи для завантаження даних, їх обробки, візуалізації та генерації звітів. Така структура дозволяє легко масштабувати функціональність та адаптувати скрипт для аналізу різних типів експериментальних даних.

Метод __init__ відповідає за ініціалізацію об'єкта, а метод load_weights_data виконує завантаження експериментальних даних з бінарного файлу best_weights.bin, який містить збережені ваги навченої нейромережі. На відміну від традиційного CSV-файлу з текстовими даними, бінарний файл містить послідовність чисел з подвійною точністю, що представляють усі вагові коефіцієнти нейромережі: матриці ваг w_1, w_2, w_3 та вектори зміщень b_1, b_2, b_3.

Для коректного читання файлу скрипт використовує бібліотеку numpy.frombuffer, яка дозволяє інтерпретувати сирі байти як масив чисел з плаваючою комою [14]. Після завантаження ваги конвертуються у масив NumPy для подальших обчислень. Важливою особливістю є автоматичне визначення розміру нейромережі на основі структури, закладеної в контролері: вхідний шар (3 нейрони), перший прихований шар (6 нейронів), другий прихований шар (12 нейронів) та вихідний шар (3 нейрони) (рисунок 6).

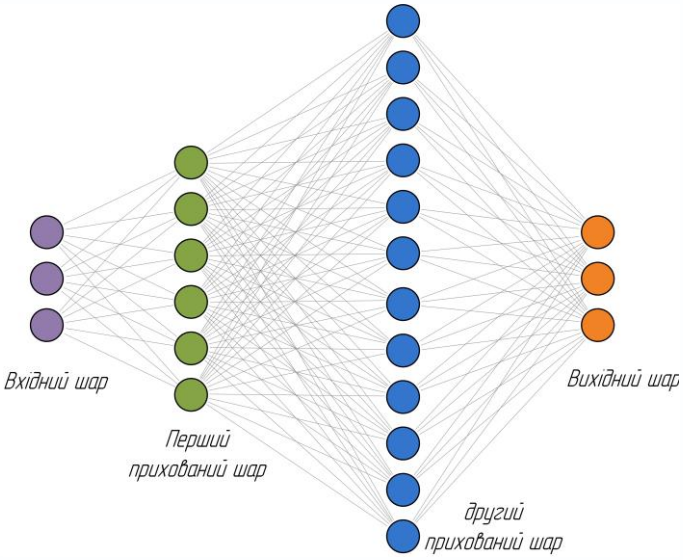
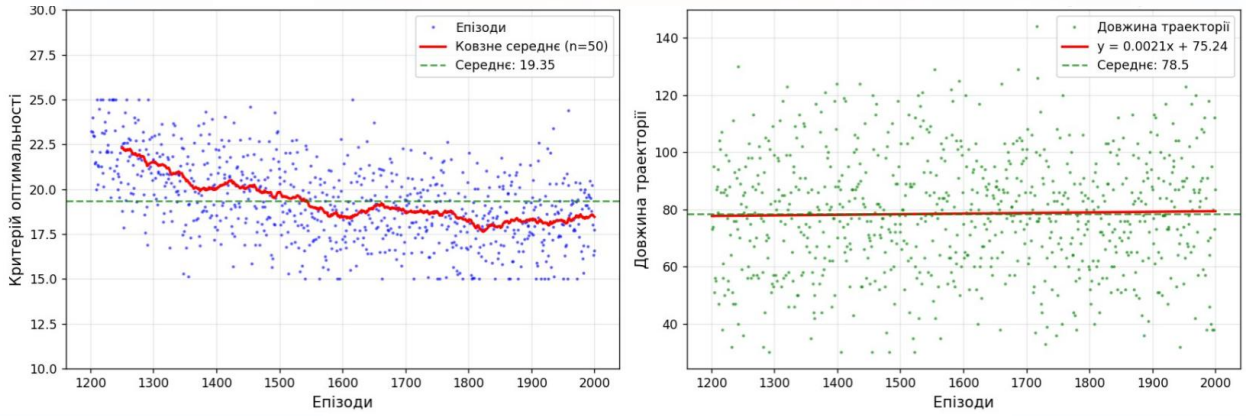


Рисунок 6 – Структура нейронної мережі

Це дозволяє правильно інтерпретувати послідовність ваг та відновити повну топологію нейромережі.

На рисунку 7 представлені результати навчання на фінальному етапі (епізоди 1200-2000). На графіку а спостерігається суттєва зміна характеру поведінки робота: критерій оптимальності стабілізується у вузькому діапазоні 18-22, що підтверджує формування стабільної політики управління.

ISSN 2786-6025 Online

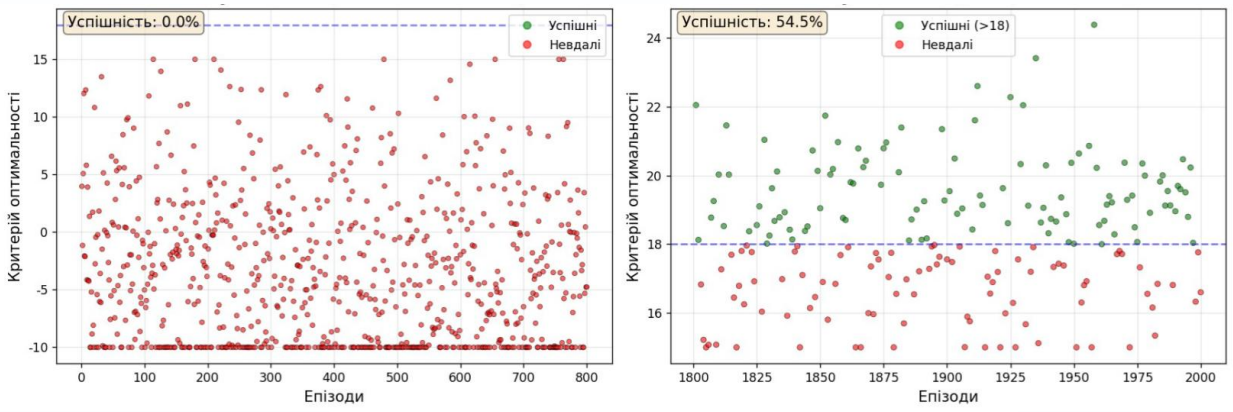


а) б)

Рисунок 7 – Результати навчання:
а) стабільна політика; б) оптимальні траєкторії.

Червона лінія ковзного середнього практично збігається із зеленим пунктиром, що показує середнє значення критерію, а незначні коливання пояснюються випадковими факторами середовища. Графік б демонструє оптимальні траєкторії руху: довжина траєкторій значно скоротилася до середнього значення 78,5 кроків. Важливо відзначити, що розкид значень мінімальний, що свідчить про високу відтворюваність результатів. Лінія тренду має незначний негативний нахил, що вказує на подальше невелике покращення ефективності руху. Зелений пунктир показує середнє значення довжини траєкторії, яке є близьким до теоретично оптимального для заданої конфігурації середовища. Така поведінка свідчить про успішне завершення процесу навчання та формування оптимальної стратегії навігації.

На рисунку 8, а представлено порівняння успішності робота на початковому етапі навчання, а на рисунку 8, б – на фінальному. Лівий графік відображає розподіл епізодів на початковому етапі навчання, де зелені точки позначають успішні епізоди (критерій оптимальності > 18), а червоні – неуспішні.



а) б)

Рисунок 8 – Динаміка навчання: а – початок навчання; б- кінець навчання

ISSN 2786-6025 Online

На цьому етапі переважають червоні точки, що підтверджує низьку ефективність агента. Синя горизонтальна лінія позначає поріг успішності (критерій=18), вище якого епізод вважається успішним. Правий графік демонструє кардинально іншу картину для фінального етапу: переважна більшість точок є зеленими, що свідчить про досягнення високої успішності на рівні 54.5%.

На рисунку 9 представлено комплексний аналіз динаміки навчання протягом всіх 2000 епізодів. Основна синя лінія показує значення критерію оптимальності для кожного епізоду, а червона лінія відображає ковзне середнє з вікном 50 епізодів, що дозволяє виявити загальний тренд, згладжуючи випадкові коливання.

Зелена штрихова лінія демонструє найкращий досягнутий результат на кожному етапі навчання, який монотонно зростає від -5 до 25. Сіра зона (епізоди 1-800) – період стохастичної поведінки з високою варіативністю; жовта зона (епізоди 800-1200) – перехідний період, де відбувається формування стабільної політики; зелена зона (епізоди 1200-2000) – період стабільної роботи з високими та стабільними значеннями критерію.

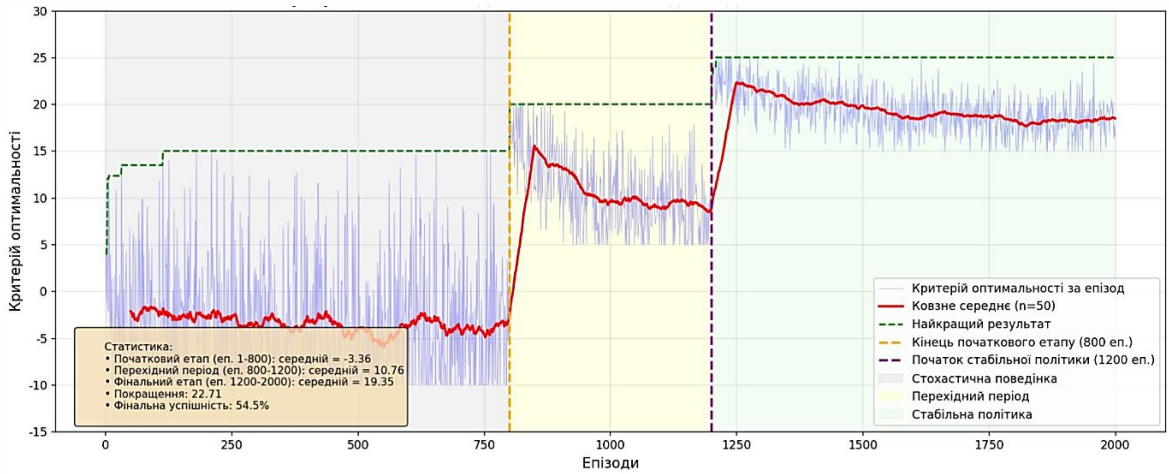
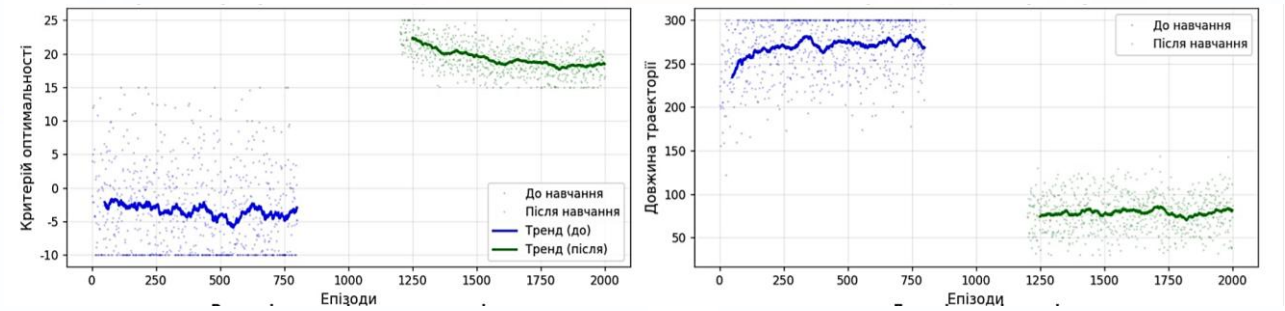


Рисунок 9 – Прогрес навчання протягом всіх епізодів

На рисунку 10 представлено графіки, що дозволяють провести всебічний порівняльний аналіз результатів до та після навчання. Графік ліворуч показує гістограми розподілу критерію оптимальності: синій стовпчик відповідає початковому етапу (епізоди 1-800) і має широкий розподіл з центром у діапазоні 0-5, зелений стовпчик – фінальному етапу (епізоди 1200-2000) з вузьким розподілом у діапазоні 18-22. Графік праворуч ілюструє аналогічне порівняння для довжини траєкторій: спостерігається суттєве зміщення розподілу від довгих траєкторій (200-300 кроків) до коротких оптимальних (50-80 кроків).

ISSN 2786-6025 Online



а) б)

Рисунок 10 – Порівняльний аналіз результатів навчання:
а – критерію оптимальності; б – довжини траєкторії

На рисунку 11 представлено результати аналізу довжини траєкторії руху агента залежно від номера епізоду.

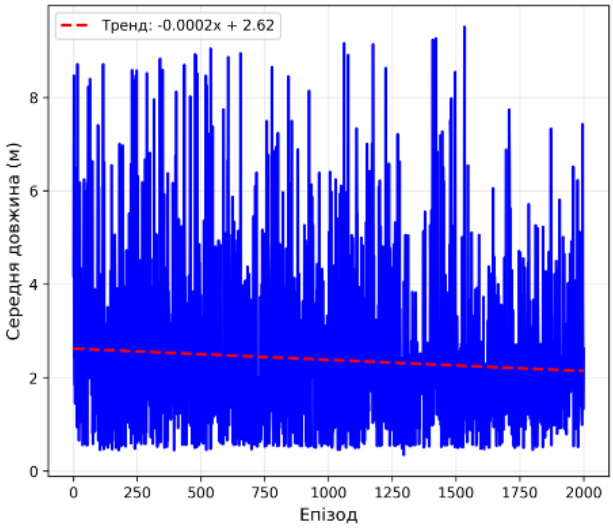


Рисунок 11 – Результати аналізу довжини траєкторії руху агента

Аналіз граничних значень показує, що мінімальна довжина траєкторії зменшується від 4,5 м на початковому етапі до 1,8 м наприкінці навчання. Це значення наближається до оптимальної довжини траєкторії (1,5 м) з урахуванням необхідності обходу перешкоди. Водночас максимальна довжина траєкторії зменшується від 17,5 м до 8,2 м, що свідчить про поступову стабілізацію поведінки робота. Після завершення навчання проведено тестування у режимі експлуатації (exploitation) з використанням зафіксованих найкращих ваг мережі.

Результати експериментів показали, що робот досягав цілі у понад 54% випадків, успішно уникаючи перешкод та формуючи траєкторії, близькі до оптимальних, без застосування зовнішніх правил чи евристичних інструкцій.

ISSN 2786-6025 Online

Висновки. У межах виконаної роботи створено та експериментально перевірено програмний модуль реалізації алгоритму навчання з підкріпленням для керування мобільним роботом у середовищі Webots з функцією адаптивного огинання перешкод.

У результаті навчання тривалістю 2000 епізодів критерій оптимальності зріс з середнього значення -3,36 на початковому етапі (епізоди 1-800) до 10,76 у перехідному періоді (епізоди 800-1200) та до 19,35 на фінальному етапі (епізоди 1200-2000). Загальне покращення показника склало 22,71, що підтверджує ефективність запропонованого методу навчання.

Фінальна успішність досягнення цільової точки на завершальному етапі навчання склала 54,5%, що демонструє здатність агента до стабільного виконання поставленого завдання.)

Залучення платформи Webots як середовища моделювання надало змогу реалізувати повноцінний навчальний процес у режимі віртуального часу, забезпечивши прискорення обчислень. Інтеграція з фізичним рушієм ODE гарантує достовірність отриманих даних і створює передумови для успішного перенесення навченої моделі на апаратні платформи.

Література:

1. Свідоцтво про державну реєстрацію авторського права на твір. Комп'ютерна програма «Контролер керування мобільним роботом в середовищі Webots на основі алгоритму навчання з підкріпленням» / Р. С. Белзецький, Є. Р. Добровольська, А. В. Чернокнижник. – №145166. – Україна, 2026.

2. Rybczak, M., Popowniak, N., Lazarowska, A. A Survey of Machine Learning Approaches for Mobile Robot Control. [Електронний ресурс] / M. Rybczak, N. Popowniak, A. Lazarowska // Robotics. – 2024. 1, № 13. – Режим доступу: https://doi.org/10.3390/ robotics 13010012. (дата звернення: 8.04.2026).

3. Pfeiffer, M., Schaeuble, M., Nieto, J., Siegwart, R., Cadena, C. From perception to decision: A data-driven approach to end-to-end motion planning for autonomous ground robots. [Електронний ресурс] / M. Pfeiffer, M. Schaeuble, J. Nieto, R. Siegwart, C. Cadena // 2017 IEEE International Conference on Robotics and Automation (ICRA). – 2017. – Режим доступу: 10.1109/ICRA.2017.7989182. (дата звернення: 8.04.2026).

4. Tai, L., Paolo, G., Liu, M. Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation. [Електронний ресурс] / L. Tai, G. Paolo, M. Liu. // Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International. – 2017. – Режим доступу: 10.1109/IROS.2017.8202134. (дата звернення: 8.04.2026).

5. Faust, A., Ramirez, O., Fiser, M., Oslund K., Francis A., Davidson, J., and Tapia, L. PRM-RL: Long-range Robotic Navigation Tasks by Combining Reinforcement Learning and Sampling-based Planning. [Електронний ресурс] / A. Faust, O. Ramirez, M. Fiser, K. Oslund, A. Francis, J. Davidson, and L. Tapia. // IEEE International Conference on Robotics and Automation (ICRA). – 2017. – Режим доступу: https://doi.org/10.48550/arXiv.1710.0393. (дата звернення: 8.04.2026).

6. Francis, G., Faust, A., and Tapia, L. Long-Range Indoor Navigation With PRM-RL. [Електронний ресурс] / G. Francis, A. Faust, and L. Tapia. // IEEE Transactions on Robotics. – 2020. – Режим доступу: 10.1109/TRO.2020.2975428. (дата звернення: 8.04.2026).

ISSN 2786-6025 Online

7. Michel, O. Cyberbotics Ltd. WebotsTM: Professional Mobile Robot Simulation. [Електронний ресурс] / O. Michel. // International Journal of Advanced Robotic Systems. – 2004. – Режим доступу: 10.5772/5618. (дата звернення: 8.04.2026).

8. Створення віртуального прототипу мобільного робота в середовищі Webots. Белзецький Р.С., Чернокнижник А. В., Добровольська Є. Р., Чіпак О. С. Матеріали Всеукраїнська науково-технічна конференція підрозділів ВНТУ (ВНТКП ВНТУ) ВНТУ, Вінниця, 24-27 березня 2026 р. Електрон. текст. дані. 2026. URI: https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2026/paper/view (дата звернення: 8.04.2026).

9. Fillion-Robin, J. Modeling of a real quadruped robot using Webots(TM) simulation platform. [Електронний ресурс] / J. Fillion-Robin. // – 2007. – Режим доступу: https://www.researchgate.net/publication/260683918_Modeling_of_a_real_quadruped_robot_using_WebotsTM_simulation_platform.

10. Webots Reference Manual [Електронний ресурс]. – Режим доступу: https://cyberbotics.com/doc/reference/index. (дата звернення: 8.04.2026).

11. Garcia, F., Rachelson, E. Markov decision processes: підручник [Електронний ресурс] / F. Garcia, E. Rachelson. – Markov Decision Processes in Artificial Intelligence, 2013. – 38 с. – Режим доступу: https://doi.org/10.1002/9781118557426.ch1.

12. Pan, J., Wang, X., Cheng, Y., Yu, Q. Multisource transfer double DQN based on actor learning. [Електронний ресурс] / J. Pan, X. Wang, Y. Cheng, Q. Yu. // IEEE Transactions on Neural Networks and Learning Systems. – 2018. – Режим доступу: 10.1109/TNNLS.2018.2806087. (дата звернення: 13.04.2026).

13. LI, Zhengyang. Advances in Deep Reinforcement Learning for Computer Vision Applications. [Електронний ресурс] / Zhengyang LI. // Journal of Industrial Engineering and Applied Science. – 2024. – Режим доступу: 10.70393/6a69656173.323234. (дата звернення: 13.04.2026).

14. NumPy. – Режим доступу: [Електронний ресурс]. – Режим доступу: https://numpy.org/doc/stable/reference/generated/numpy.frombuffer.html (дата звернення: 13.03.2026).

References:

1. Svidotstvo pro derzhavnu reiestratsiiu avtorskoho prava na tvir. Kompiuterna prohrama «Kontroler keruvannia mobilnym robotom v seredovyshchi Webots na osnovi alhorytmu navchannia z pidkriplenniam» / R. S. Belzetskyi, Ye. R. Dobrovolska, A. V. Chernoknyzhnyk. – №145166. – Ukraina, 2026.

2. Rybczak, M., Popowniak, N., Lazarowska, A. A Survey of Machine Learning Approaches for Mobile Robot Control. [Elektronnyi resurs] / M. Rybczak, N. Popowniak, A. Lazarowska // Robotics. – 2024. 1, № 13. – Rezhym dostupu: https://doi.org/10.3390/ robotics 13010012. (data zvernennia: 8.04.2026).

3. Pfeiffer, M., Schaeuble, M., Nieto, J., Siegwart, R., Cadena, C. From perception to decision: A data-driven approach to end-to-end motion planning for autonomous ground robots. [Elektronnyi resurs] / M. Pfeiffer, M. Schaeuble, J. Nieto, R. Siegwart, C. Cadena // 2017 IEEE International Conference on Robotics and Automation (ICRA). – 2017. – Rezhym dostupu: 10.1109/ICRA.2017.7989182. (data zvernennia: 8.04.2026).

4. Tai, L., Paolo, G., Liu, M. Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation. [Elektronnyi resurs] / L. Tai, G. Paolo, M. Liu. // Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International. – 2017. – Rezhym dostupu: 10.1109/IROS.2017.8202134. (data zvernennia: 8.04.2026).

ISSN 2786-6025 Online

5. Faust, A., Ramirez, O., Fiser, M., Oslund K., Francis A., Davidson, J., and Tapia, L. PRM-RL: Long-range Robotic Navigation Tasks by Combining Reinforcement Learning and Sampling-based Planning. [Elektronnyi resurs] / A. Faust, O. Ramirez, M. Fiser, K. Oslund, A. Francis, J. Davidson, and L. Tapia. // IEEE International Conference on Robotics and Automation (ICRA). – 2017. – Rezhym dostupu: https://doi.org/10.48550/arXiv.1710.0393. (data zvernennia: 8.04.2026).

6. Francis, G., Faust, A., and Tapia, L. Long-Range Indoor Navigation With PRM-RL. [Elektronnyi resurs] / G. Francis, A. Faust, and L. Tapia. // IEEE Transactions on Robotics. – 2020. – Rezhym dostupu: 10.1109/TRO.2020.2975428. (data zvernennia: 8.04.2026).

7. Michel, O. Cyberbotics Ltd. WebotsTM: Professional Mobile Robot Simulation. [Elektronnyi resurs] / O. Michel. // International Journal of Advanced Robotic Systems. – 2004. – Rezhym dostupu: 10.5772/5618. (data zvernennia: 8.04.2026).

8. Stvorennia virtualnoho prototypu mobilnoho robota v seredovyshchi Webots. Belzetskyi R.S., Chernoknyzhnyk A. V., Dobrovolska Ye. R., Chipak O. S. Materialy Vseukrainska naukovo-tekhnichna konferentsiia pidrozdiliv VNTU (VNTKP VNTU) VNTU, Vinnytsia, 24-27 bereznia 2026 r. Elektron. tekst. dani. 2026. URI: https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2026/paper/view (data zvernennia: 8.04.2026).

9. Fillion-Robin, J. Modeling of a real quadruped robot using Webots(TM) simulation platform. [Elektronnyi resurs] / J. Fillion-Robin. // – 2007. – Rezhym dostupu: https://www.researchgate.net/publication/260683918_Modeling_of_a_real_quadruped_robot_using_WebotsTM_simulation_platform.

10. Webots Reference Manual [Elektronnyi resurs]. – Rezhym dostupu: https://cyberbotics.com/doc/reference/index. (data zvernennia: 8.04.2026).

11. Garcia, F., Rachelson, E. Markov decision processes: pidruchnyk [Elektronnyi resurs] / F. Garcia, E. Rachelson. – Markov Decision Processes in Artificial Intelligence, 2013. – 38 s. – Rezhym dostupu: https://doi.org/10.1002/9781118557426.ch1.

12. Pan, J., Wang, X., Cheng, Y., Yu, Q. Multisource transfer double DQN based on actor learning. [Elektronnyi resurs] / J. Pan, X. Wang, Y. Cheng, Q. Yu. // IEEE Transactions on Neural Networks and Learning Systems. – 2018. – Rezhym dostupu: 10.1109/TNNLS.2018.2806087. (data zvernennia: 13.04.2026).

13. LI, Zhengyang. Advances in Deep Reinforcement Learning for Computer Vision Applications. [Elektronnyi resurs] / Zhengyang LI. // Journal of Industrial Engineering and Applied Science. – 2024. – Rezhym dostupu: 10.70393/6a69656173.323234. (data zvernennia: 13.04.2026).

14. NumPy. – Rezhym dostupu: [Elektronnyi resurs]. – Rezhym dostupu: https://numpy.org/doc/stable/reference/generated/numpy.frombuffer.html (data zvernennia: 13.03.2026).

Дата першого надходження статті до видання: 11.04.2026
Дата прийняття статті до друку після рецензування: 26.04.2026