

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ
ТЕХНОЛОГІЙ

Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

ГАНЕНКО ЛЮДМИЛА ДМИТРІВНА

УДК 004.85:007.52(043)

ДИСЕРТАЦІЯ
МЕТОДИ ТА МОДЕЛЬ ІНТЕЛЕКТУАЛЬНОЇ НАВІГАЦІЇ АВТОНОМНИХ
МОБІЛЬНИХ РОБОТІВ У ДИНАМІЧНОМУ СЕРЕДОВИЩІ НА ОСНОВІ
ГЛИБИННОГО НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Спеціальність 123 «Комп'ютерна інженерія»

Подається на здобуття наукового ступеня
доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

_____ Л. Д. Ганенко

Науковий керівник:
Аронов Андрій Олексійович
кандидат технічних наук

Київ 2026

АНОТАЦІЯ

Ганенко Л. Д. Методи та модель інтелектуальної навігації автономних мобільних роботів у динамічному середовищі на основі глибинного навчання з підкріпленням. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії з галузі знань 12 «Інформаційні технології» за спеціальністю 123 «Комп'ютерна інженерія» – Державний університет інформаційно-комунікаційних технологій Міністерства освіти і науки України, Київ, 2026.

Сучасний етап розвитку робототехніки визначається зміною парадигми від експлуатації роботів у детермінованих промислових зонах до їхньої повноцінної інтеграції в неструктуроване соціальне середовище. Мобільні сервісні платформи, призначені для роботи в офісах, логістичних центрах або медичних закладах, функціонують в динамічному середовищі із присутністю людей.

На відміну від статичних перешкод, люди є активними агентами зі стохастичною поведінкою. Їхнє переміщення підпорядковується не лише фізичним законам, а й соціальним нормам – правилам етикету та груповій динаміці. Ігнорування цих чинників у навігаційних моделях призводить до деструктивної взаємодії, за якої поведінка робота сприймається оточенням як інтуїтивно незрозуміла або потенційно загрозлива.

Класичні методи навігації демонструють обмежену ефективність у соціальному динамічному середовищі. Вони інтерпретують людей як динамічні об'єкти, не враховуючи специфіку соціального простору.

Застосування методів глибинного навчання з підкріпленням (Deep Reinforcement Learning, DRL) дозволяє перейти від програмування правил до формування оптимальних стратегій поведінки через взаємодію із середовищем. DRL-агенти здатні виявляти неявні закономірності в людській поведінці. Проте існуючі алгоритми мають обмеження, зокрема розрідженість функцій винагороди, тривалу збіжність процесу навчання, недостатню стійкість до стохастичної поведінки людей. Тому розробка методів інтелектуальної навігації, які поєднують

DRL із механізмами прогнозування невизначеності та адаптивного навчання, є актуальним завданням наукових досліджень.

Метою дослідження є підвищення ефективності та безпеки навігації автономних мобільних роботів у динамічних соціальних середовищах шляхом розробки моделей та методів на основі глибинного навчання з підкріпленням.

Об'єкт дослідження – процес інтелектуальної навігації автономних мобільних роботів у динамічних середовищах із присутністю людей.

Предмет дослідження – моделі та методи глибинного навчання з підкріпленням для забезпечення інтелектуальної навігації автономних мобільних роботів.

Методи дослідження. Для вирішення поставлених завдань у роботі використано: методи теорії автоматичного керування та автономної навігації – для формування алгоритмів руху робота; методи глибинного навчання з підкріпленням – для формування інтелектуальної стратегії навігації; методи теорії ймовірностей та математичної статистики – для обробки стохастичних даних та оцінки невизначеності середовища; принципи об'єктно-орієнтованого програмування та імітаційного моделювання – для програмної реалізації та експериментальної апробації системи керування.

Наукова новизна одержаних результатів полягає у наступному:

1. Удосконалено модель соціально-адаптивної навігації автономного мобільного робота на основі формалізації марковського процесу прийняття рішень, яка, на відміну від існуючих, завдяки інтеграції кінематичних обмежень робота з параметрами соціальної взаємодії, дозволяє системі керування ідентифікувати та класифікувати потенційні проксемічні конфлікти в режимі реального часу й підвищити безпеку руху в динамічному соціальному середовищі.

2. Вперше розроблено метод навчання навігаційної політики на основі глибинного навчання з підкріпленням, який за рахунок застосування комбінованої стратегії Curriculum Learning із механізмом автоматизованого переходу між етапами складності дозволяє вирішити проблему розрідженої винагороди та

забезпечити прискорення збіжності нейромережевої моделі в умовах соціальної навігації.

3. Вперше розроблено метод адаптивного формування винагороди на основі прогнозу невизначеності динамічного середовища, який за рахунок впровадження механізму динамічного зважування компонентів функції винагороди дозволяє оптимізувати стратегію поведінки автономного мобільного робота залежно від рівня невизначеності середовища.

Практичне значення дисертаційного дослідження. Удосконалена модель соціально-адаптивної навігації автономного мобільного робота, яка завдяки інтеграції кількісних параметрів проксеміки трансформує неявні правила поведінки людей у алгоритмічні обмеження, дозволяє генерувати передбачувані безконфліктні траєкторії руху в соціальному середовищі. Розроблений метод навчання навігаційної політики на основі глибинного навчання з підкріпленням та Curriculum Learning з автоматизованим механізмом переходу між етапами складності, який шляхом послідовної декомпозиції навігаційних завдань забезпечує стабільну збіжність політики керування та зростання частоти досягнення автономним мобільним роботом заданої цілі із одночасним зменшенням кількості фізичних зіткнень. Розроблений метод адаптивного формування винагороди на основі прогнозу невизначеності динамічного середовища, який з урахуванням поточного рівня невизначеності автоматично коригує пріоритети між цільовою ефективністю та соціальною безпекою, надає змогу програмному забезпеченню сервісних мобільних систем генерувати безпечні стратегії руху в стохастичних умовах.

У першому розділі проведено комплексний аналіз проблеми автономної навігації у середовищах із присутністю людей. Визначено, що ключовим викликом для комп'ютерних систем керування є перехід від фізичного уникнення перешкод до забезпечення соціальної передбачуваності маневрів. Встановлено обмеженість класичних методів через їхній реактивний характер та нездатність враховувати часові залежності у поведінці людей.

В другому розділі формалізовано модель соціально-адаптивної навігації автономного мобільного робота, в якій задачу керування представлено як марковський процес прийняття рішень із розширеним простором станів. На відміну від класичних моделей, запропонований підхід інтегрує кількісні параметри проксеміки за Е. Холлом безпосередньо у функцію винагороди. Це дозволило трансформувати неявні соціальні норми у чіткі математичні обмеження, забезпечуючи навчання агента діяти не лише безпечно, а й передбачувано для людей.

Розроблено метод PPO-CL навчання автономного мобільного робота на основі глибинного навчання з підкріпленням та стратегії Curriculum Learning, який передбачає ієрархічну декомпозицію задачі навігації на чотири послідовні етапи. Запропоновано автоматизований механізм переходу між етапами навчання на основі аналізу стаціонарності, стабільності та успішності процесу навчання.

У третьому розділі спроектовано та обґрунтовано архітектуру інтелектуального агента на основі алгоритму PPO, інтегрованого з гібридним модулем імовірнісного прогнозування LSTM-MDN. Використання MDN у поєднанні з рекурентними шарами дозволило апроксимувати мультимодальний розподіл імовірностей майбутніх положень людей. Це дало змогу кількісно оцінити міру невизначеності середовища та перейти від реактивного до проактивного планування маневрів.

Удосконалено механізм формування функції винагороди через впровадження динамічного зважування компонентів. Розроблений метод дозволив забезпечити адаптивне коригування пріоритетів між швидкістю досягнення цілі та дотриманням соціального комфорту залежно від рівня невизначеності. У ситуаціях із високою мірою невизначеності робот автоматично надає перевагу стратегії, яка мінімізує ризик виникнення небезпечних ситуацій.

Четвертий розділ присвячено експериментальній апробації. Створено спеціалізоване симуляційне середовище на основі Gymnasium та моделі соціальних сил. Для подолання розриву між симуляцією та реальністю у моделі враховано стохастичні похибки сенсорних систем та латентність обчислювальних процесів.

Експериментально підтверджено, що запропонований метод PPO-CL порівняно з базовим PPO дозволяє підвищити успішність виконання завдань та зменшити кількість критичних зіткнень.

Для реалізації проактивної поведінки робота в умовах невизначеності динамічного середовища досліджено гібридну архітектуру інтелектуального агента, інтегровану з модулем імовірнісного прогнозування на основі рекурентних нейронних мереж (PPO-LSTM-MDN). Розроблена система демонструє здатність до проактивного формування траєкторій, які характеризуються мінімальними часовими витратами при максимальному рівні безпеки та соціальної прийнятності поведінки робота.

За результатами дисертаційних досліджень опубліковано 17 наукових праць. Основний зміст дослідження відображено у 6 наукових статтях, опублікованих у фахових виданнях, які затверджено наказом МОН України. Матеріали дисертації представлено на 11 науково-практичних конференціях і опубліковано у збірниках тез доповідей, з яких 1 публікація у виданні, індексованому у міжнародній наукометричній базі Scopus.

У дисертаційній роботі вирішено науково-практичне завдання підвищення ефективності інтелектуальної навігації автономних мобільних роботів у динамічному соціальному середовищі. Запропоновані методи дозволяють автономним мобільним роботам не лише уникати фізичних зіткнень, а й дотримуватися соціальних норм для їх безпечної експлуатації.

Ключові слова: інформаційні технології, машинне навчання, штучний інтелект, інтелектуальна навігація, соціальна навігація, автономні мобільні роботи, робототехніка, навчання з підкріпленням, глибинне навчання, глибинне навчання з підкріпленням, нейронні мережі, модель, кіберфізичні системи, програмно-апаратні засоби, комп'ютерні системи реального часу, інтелектуальні комп'ютерні системи.

ABSTRACT

Hanenko L. D. Methods and a model for the intelligent navigation of autonomous mobile robots in a dynamic environment based on deep reinforcement learning. – Qualification scientific work in the form of a manuscript.

Thesis for the degree of Doctor of Philosophy in the field of knowledge 12 “Information Technologies” in the specialty 123 “Computer Engineering” – State University of Information and Communication Technologies of the Ministry of Education and Science of Ukraine, Kyiv, 2026.

The current stage of robotics development is characterised by a paradigm shift from the operation of robots in deterministic industrial zones to their full integration into an unstructured social environment. Mobile service platforms designed for use in offices, logistics centres or healthcare facilities operate in a dynamic environment where people are present.

Unlike static obstacles, people are active agents with stochastic behaviour. Their movement is governed not only by physical laws but also by social norms – rules of etiquette and group dynamics. Ignoring these factors in navigation models leads to destructive interactions, in which the robot’s behaviour is perceived by its surroundings as intuitively incomprehensible or potentially threatening.

Traditional navigation methods demonstrate limited effectiveness in socially dynamic environments. They treat people as dynamic objects without taking into account the specific nature of social space.

The application of deep reinforcement learning (DRL) methods enables a shift from rule-based programming to the formation of optimal behavioural strategies through interaction with the environment. DRL agents are capable of identifying implicit patterns in human behaviour. However, existing algorithms have limitations, including sparse reward functions, slow convergence of the learning process, and insufficient robustness to stochastic human behaviour. Therefore, the development of intelligent navigation methods that combine DRL with mechanisms for uncertainty prediction and adaptive learning is a pressing research priority.

The aim of this research is to improve the efficiency and safety of autonomous mobile robot navigation in dynamic social environments by developing models and methods based on reinforcement learning.

The object of the study is the process of intelligent navigation of autonomous mobile robots in dynamic environments where people are present.

The subject of the study is deep reinforcement learning models and methods for ensuring the intelligent navigation of autonomous mobile robots.

Research methods. To address the objectives set out in this study, the following methods were employed: methods of automatic control theory and autonomous navigation – for developing the robot’s motion algorithms; reinforcement learning methods – for developing an intelligent navigation strategy; methods of probability theory and mathematical statistics – for processing stochastic data and assessing environmental uncertainty; principles of object-oriented programming and simulation modelling – for the software implementation and experimental testing of the control system.

The scientific novelty of the results obtained lies in the following:

1. The model of socially adaptive navigation for an autonomous mobile robot has been improved based on the formalisation of a Markov decision process which, unlike existing models, thanks to the integration of the robot’s kinematic constraints with social interaction parameters, enables the control system to identify and classify potential proxemic conflicts in real time and enhance traffic safety in a dynamic social environment.

2. For the first time, a method has been developed for training navigation policies based on deep reinforcement learning, which, through the application of a combined Curriculum Learning strategy with a mechanism for automated transition between levels of complexity, enables the resolution of the sparse reward problem and ensures accelerated convergence of the neural network model in social navigation conditions.

3. For the first time, a method has been developed for the adaptive formulation of rewards based on a forecast of the uncertainty in a dynamic environment; by introducing a mechanism for the dynamic weighting of the components of the reward

function, this method enables the optimisation of an autonomous mobile robot's behavioural strategy depending on the level of uncertainty in the environment.

Practical significance of the thesis research. The improved model of socially adaptive navigation for an autonomous mobile robot, which—by integrating quantitative parameters of proxemics—transforms implicit rules of human behaviour into algorithmic constraints, enables the generation of predictable, conflict-free movement trajectories in a social environment. A method has been developed for training navigation policies based on deep reinforcement learning and Curriculum Learning with an automated mechanism for transitioning between levels of complexity, which, through the sequential decomposition of navigation tasks, ensures stable convergence of the control policy and an increase in the frequency with which an autonomous mobile robot achieves a given goal, whilst simultaneously reducing the number of physical collisions. A method has been developed for adaptive reward formation based on the prediction of uncertainty in a dynamic environment, which, taking into account the current level of uncertainty, automatically adjusts the priorities between target efficiency and social safety, enabling the software of service mobile systems to generate safe movement strategies under stochastic conditions.

The first chapter presents a comprehensive analysis of the problem of autonomous navigation in environments where people are present. It is established that the key challenge for computer control systems is the transition from physical obstacle avoidance to ensuring the social predictability of manoeuvres. The limitations of classical methods are established due to their reactive nature and inability to account for temporal dependencies in human behaviour.

The second chapter formalises a model of socially adaptive navigation for an autonomous mobile robot, in which the control problem is represented as a Markov decision process with an extended state space. Unlike classical models, the proposed approach integrates the quantitative parameters of proxemics according to E. Hall directly into the reward function. This has made it possible to transform implicit social norms into clear mathematical constraints, ensuring that the agent learns to act not only safely but also predictably for people.

A PPO-CL method has been developed for training an autonomous mobile robot based on deep reinforcement learning and the Curriculum Learning strategy, which involves a hierarchical decomposition of the navigation task into four sequential stages. An automated mechanism for transitioning between learning stages has been proposed, based on an analysis of the stationarity, stability and success of the learning process.

In the third chapter, the architecture of an intelligent agent based on the PPO algorithm, integrated with a hybrid LSTM-MDN probabilistic forecasting module, is designed and justified. The use of mixed-density networks in combination with recurrent layers made it possible to approximate the multimodal probability distribution of people's future positions. This allowed for a quantitative assessment of the degree of environmental uncertainty and a shift from reactive to proactive manoeuvre planning.

The mechanism for generating the reward function has been improved through the introduction of dynamic weighting of components. The developed method has enabled adaptive adjustment of priorities between the speed of achieving the goal and maintaining social comfort, depending on the level of uncertainty. In situations with a high degree of uncertainty, the robot automatically favours a strategy that minimises the risk of dangerous situations arising.

The fourth chapter is devoted to experimental validation. A specialised simulation environment has been created based on Gymnasium and a model of social forces. To bridge the gap between simulation and reality, the model accounts for stochastic errors in sensor systems and the latency of computational processes.

It has been experimentally confirmed that, compared to the baseline PPO, the proposed PPO-CL method improves task success rates and reduces the number of critical collisions.

To implement proactive behaviour in a dynamic, uncertain environment, a hybrid intelligent agent architecture has been investigated, integrated with a probabilistic forecasting module based on recurrent neural networks and mixed-density networks (PPO-LSTM-MDN). The developed system demonstrates the ability to proactively generate trajectories that minimise travel time whilst ensuring maximum safety and comfort for those around it.

Seventeen academic papers have been published based on the findings of the dissertation research. The main content of the research is reflected in six academic articles published in specialist journals approved by order of the Ministry of Education and Science of Ukraine. The materials of the thesis have been presented at 11 scientific and practical conferences and published in proceedings, including one publication in a journal indexed in the international scientometric database Scopus.

The thesis addresses the scientific and practical task of improving the efficiency of intelligent navigation of autonomous mobile robots in a dynamic social environment. The proposed methods enable autonomous mobile robots not only to avoid physical collisions but also to adhere to social norms for their safe operation.

Keywords: information technology, machine learning, artificial intelligence, intelligent navigation, social navigation, autonomous mobile robots, robotics, reinforcement learning, deep learning, deep reinforcement learning, neural networks, models, cyber-physical systems, hardware and software, real-time computer systems, intelligent computer systems.

Список опублікованих праць за темою дисертації

Наукові праці, в яких опубліковані основні наукові результати дисертації:

1. Ганенко Л. Д., Жебка В. В. Аналітичний огляд питань навігації мобільних роботів в закритих приміщеннях. *Телекомунікаційні та інформаційні технології*. 2023. №3. С. 85-96.
2. Ганенко Л. Д., Жебка В. В. Застосування методів навчання з підкріпленням для планування шляху мобільних роботів. *Телекомунікаційні та інформаційні технології*. 2024. №1. С. 16-25.
3. Ганенко Л. Д., Жебка В. В. Створення навігаційної системи автономного мобільного робота засобами ROS 2. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*. 2025 №4(28), С.498–510.
4. Ганенко Л. Д., Жебка В. В. Модель соціально-адаптивної навігації мобільного робота з використанням методів навчання з підкріпленням. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*. 2025 №1 (29), С. 559-570.
5. Ганенко Л. Д., Бушма О. В. Метод навчання автономних мобільних роботів на основі DRL та Curriculum Learning. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*. 2025. № 2(30), С. 568–582.
6. Ганенко Л. Д. Метод адаптивного формування винагороди за умов невизначеності динамічних об'єктів. *Телекомунікаційні та інформаційні технології*. 2026. №1. С. 23-30.

Наукові праці, які засвідчують апробацію матеріалів дисертації:

7. Ганенко Л. Д., Жебка В. В. Особливості планування шляху мобільного робота. *Технологічні горизонти: дослідження та застосування інформаційних технологій для технологічного прогресу України та світу: зб. тез Всеукр. наук.-техн. конф., м. Київ, 28 листопада 2023 р. Київ: ДУІКТ, 2023. С. 193-195.*

8. Ганенко Л. Д. Інтелектуальні методи навігації мобільних роботів на основі SLAM: виклики та перспективи. *Цифрова гуманістика: Інформаційні технології та інформаційне моделювання на сучасному етапі розвитку суспільства*: зб. тез Всеукр. наук.-практ. конф., м. Кропивницький, 4-5 червня 2024 р. Кропивницький: ЦДУ ім. В. Винниченка, 2024. С. 186-189
9. Ганенко Л. Д., Жебка В. В. Методи навчання з підкріпленням у навігаційних системах мобільних роботів. *Інновації*: тези доп. наук. конф. молодих вчен., м. Київ, 19 вересня 2024 р. Київ: ДУІКТ, 2024. С. 28–30.
10. Ганенко Л. Д., Жебка В. В. Моделювання середовища автономних мобільних роботів. *Технологічні горизонти: дослідження та застосування інформаційних технологій для технологічного прогресу України і світу*: зб. тез II Всеукр. наук.-техн. конф., м. Київ, 18 листопада 2024 р. Київ: ДУІКТ, 2024. С. 89–90.
11. Ганенко Л. Д. Методи уникнення рухомих перешкод автономним мобільним роботом. *Проблеми комп'ютерної інженерії*: зб. тез V Всеукр. наук.-практ. конф., м. Київ, 03 грудня 2024 р. Київ: ДУІКТ, 2024. С.6–8.
12. Ганенко Л. Д., Жебка В. В. Застосування ROS для розробки робототехнічних систем. *Сучасні аспекти діджиталізації та інформатизації в програмній та комп'ютерній інженерії*: зб. тез II Міжнар. конф., м. Київ, 19-21 грудня 2024 р. Київ: ДУІКТ, 2024. С.151-153.
13. Hanenko, L., Storchak, K., Shlianchak, S., Vorokhob, M., & Pitaichuk, M. SLAM in navigation systems of autonomous mobile robots. *Cybersecurity Providing in Information and Telecommunication Systems 2025: workshop proc.*, Kyiv, 2025. Vol. 3991, P.173–182. (*Scopus*)
14. Ганенко Л. Д. Методи прогнозування руху людини в контексті безпечної взаємодії з мобільними роботами *Цифрова гуманістика: Інформаційні технології та інформаційне моделювання на сучасному етапі розвитку суспільства*: зб. тез Всеукр. наук.-практ. конф., м. Кропивницький, 22-23 травня 2025 р. Кропивницький: ЦДУ ім. В. Винниченка, 2025. С. 208–210.

15. Ганенко Л. Д., Жебка В. В. Curriculum learning як стратегія оптимізації навчання робототехнічних систем. *Виклики та рішення в програмній інженерії*: зб. тез Всеукр. наук.-тех. конф., м. Київ, 26 листопада 2025 р. Київ: ДУІКТ, 2025. С. 366–368.

16. Ганенко Л. Д., Жебка В. В. Застосування DRL для соціальної навігації автономного мобільного робота. *Проблеми комп'ютерної інженерії*: зб. тез VI Всеукр. наук.-практ. конф., м. Київ, 03 грудня 2025 р. Київ: ДУІКТ, 2025 С. 176–178.

17. Ганенко Л. Д. Метод соціально-адаптивної навігації мобільного робота. *Сучасні аспекти діджиталізації та інформатизації в програмній та комп'ютерній інженерії*: зб. тез III Міжнар. конф., м. Київ, 4-6 грудня 2025 р. Київ: ДУІКТ, 2025. С.282–284.

ЗМІСТ

ВСТУП.....	18
РОЗДІЛ 1 СУЧАСНИЙ СТАН ДОСЛІДЖЕНЬ НАВІГАЦІЇ АВТОНОМНИХ МОБІЛЬНИХ РОБОТІВ У ДИНАМІЧНОМУ СЕРЕДОВИЩІ	24
1.1. Особливості функціонування автономних мобільних роботів у динамічному середовищі.....	24
1.2. Аналіз методів планування руху мобільних роботів у динамічному середовищі.....	31
1.2.1. Методи глобального планування шляху.....	32
1.2.2. Методи локального планування траєкторії	36
1.2.3. Інтелектуальні підходи до просторової навігації та моделювання взаємодій.....	46
1.3. Постановка завдання та мети дослідження	56
Висновки до розділу 1	57
РОЗДІЛ 2 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ МОДЕЛЮВАННЯ РУХУ АВТОНОМНОГО МОБІЛЬНОГО РОБОТА.....	59
2.1. Концептуальні засади соціально-адаптивної навігації та формалізація проксемічних обмежень	59
2.2. Моделювання навігації АМР у динамічному середовищі	63
2.3. Теоретичні основи методів глибинного навчання з підкріпленням.....	64
2.3.1. Методи на основі функції цінності	67
2.3.2. Методи на основі політики.....	69
2.3.3. Методи актора–критика.....	70
2.3.4. Порівняльний аналіз алгоритмів DRL	76
2.4. Модель соціально-адаптивної навігації мобільного робота	79
2.5. Curriculum Learning як метод оптимізації навчання автономних агентів ..	86
2.6. Метод навчання навігаційної політики на основі глибинного навчання з підкріпленням та Curriculum Learning	93
Висновки до розділу 2	100
РОЗДІЛ 3 МЕТОД АДАПТИВНОГО ФОРМУВАННЯ ВИНАГОРОДИ НА ОСНОВІ ПРОГНОЗУ НЕВИЗНАЧЕНОСТІ ДИНАМІЧНОГО СЕРЕДОВИЩА ..	102
3.1. Аналіз архітектурних рішень для систем соціальної навігації в умовах невизначеності середовища.....	102

3.2. Прогнозування станів динамічних об'єктів з використанням LSTM-MDN.....	105
3.3. Модель мультимодального прогнозування станів динамічних об'єктів .	109
3.4. Метод адаптивного формування функції винагороди на основі прогнозу невизначеності динамічного середовища.....	112
Висновки до розділу 3	118
РОЗДІЛ 4 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ	120
4.1. Розробка симуляційного середовища для валідації методів соціально-адаптивної навігації	120
4.2. Система метрик оцінювання ефективності соціально-адаптивної навігації.....	126
4.3. Експериментальне дослідження методу навчання навігаційної політики на основі глибинного навчання з підкріпленням та Curriculum Learning	129
4.4. Оцінка ефективності методу адаптивного формування винагороди на основі прогнозу невизначеності.....	137
Висновки до розділу 4	143
ВИСНОВКИ	146
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	149
Додаток А. Список публікацій здобувача.....	164
Додаток Б. Акти впровадження.....	167

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

RL – Reinforcement Learning
DL – Deep Learning
DRL – Deep Reinforcement Learning
PPO – Proximal Policy Optimization
MP – мобільний робот
AMP – автономний мобільний робот
SLAM – Simultaneous Localization and Mapping
PRM – Probabilistic Roadmaps
APF – Artificial Potential Field
VFF – Virtual Force Field
VFH – Vector Field Histogram
DWA – Dynamic Window Approach
ORCA – Optimal Reciprocal Collision Avoidance
SFM – Social Force Model
MDP – Markov Decision Process
DQN – Deep Q-Network
DDQN – Double Deep Q-Network
AC – Actor-Critic
A2C – Advantage Actor-Critic
SAC – Soft Actor-Critic
A3C – Asynchronous Advantage Actor-Critic
PPO – Proximal Policy Optimization
DDPG – Deep Deterministic Policy Gradient
CL – Curriculum Learning
RNN – Recurrent Neural Network
LSTM – Long Short-Term Memory
MDN – Mixture Density Networks

ВСТУП

Актуальність теми. Сучасний етап розвитку робототехніки характеризується переходом від експлуатації роботів в ізольованих промислових зонах до їхньої інтеграції в неструктуровані соціальні середовища. Використання мобільних сервісних платформ у межах офісної чи медичної інфраструктури супроводжується постійною зміною станів середовища. У таких сценаріях людина є основним об'єктом динамічної взаємодії.

На відміну від статичних перешкод, люди є динамічними агентами зі складною стохастичною поведінкою. Рух людей підпорядковується не лише фізичним законам, а й соціальним нормам – правилам етикету та груповій динаміці. Відсутність інтеграції соціальних чинників у навігаційну модель призводить до деструктивної взаємодії, за якої поведінка робота розцінюється як інтуїтивно незрозуміла та потенційно небезпечна для людей.

Класичні методи навігації, які базуються на геометричних примітивах та локальних картах вартості, демонструють високу ефективність у статичних середовищах. Проте у динамічних соціальних середовищах вони виявляються недостатньо ефективними. Головний їхній недолік полягає у їхній реактивній природі та трактуванні людей як звичайних динамічних об'єктів.

Перспективним напрямом вирішення цієї проблеми є застосування методів глибинного навчання з підкріпленням. Підхід дозволяє перенести акцент із програмування правил на навчання агента оптимальним стратегіям поведінки шляхом спроб і помилок у середовищі. DRL-агенти здатні виявляти приховані закономірності у поведінці людей та формувати узагальнені навігаційні політики. Однак методам DRL властиві такі обмеження, як розрідженість функції винагороди та тривалий час збіжності алгоритмів.

Тому розробка методів та засобів інтелектуальної навігації, які поєднують DRL з механізмами прогнозування невизначеності та адаптивного навчання, є актуальним завданням проєктування автономних мобільних роботів, здатних функціонувати в умовах соціального динамічного середовища.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота була виконана в рамках:

1) науково-дослідної роботи «Актуальні питання сучасної інформатики та інформаційних технологій в освіті та науці» (Державний реєстраційний номер 0124U001430), Центральноукраїнського державного університету імені Володимира Винниченка;

2) госпдогвірної науково-дослідної роботи «Комплексна розробка прикладних ІТ-рішень для підвищення продуктивності комп'ютерних систем у комерційному та соціальному секторі» (Державний реєстраційний номер 0125U003178), Державного університету інформаційно-комунікаційних технологій.

Мета і завдання дослідження. Метою дослідження є підвищення ефективності та безпеки навігації автономних мобільних роботів у динамічних соціальних середовищах шляхом розробки методу адаптивного керування на основі глибинного навчання з підкріпленням.

Для досягнення поставленої мети необхідно вирішити такі завдання:

1. Проаналізувати існуючі підходи до навігації мобільних роботів у соціальному середовищі та виявити їх переваги та недоліки.
2. Розробити математичну модель соціально-адаптивної навігації на основі марковського процесу прийняття рішень (MDP) та формалізувати взаємодію робота з динамічними об'єктами.
3. Розробити метод навчання автономного мобільного робота на основі стратегії Curriculum Learning, який передбачає декомпозицію навчального процесу через поетапне ускладнення тренувальних сценаріїв у поєднанні з модифікацією структури функції винагороди.
4. Розробити архітектуру інтелектуального агента на основі алгоритму DRL для кількісної оцінки невизначеності середовища.
5. Удосконалити механізм адаптивного формування функції винагороди шляхом впровадження динамічного зважування компонентів функції винагороди.

6. Здійснити експериментальне дослідження ефективності розроблених методів у симуляційному середовищі.

Об'єкт дослідження – процес інтелектуальної навігації автономних мобільних роботів у динамічному середовищі із присутністю людей.

Предмет дослідження – моделі та методи глибинного навчання з підкріпленням для забезпечення інтелектуальної навігації автономних мобільних роботів.

Методи дослідження. Для вирішення поставлених завдань у роботі використано: методи теорії автоматичного керування та автономної навігації – для формування алгоритмів руху робота; методи глибинного навчання з підкріпленням – для формування інтелектуальної стратегії навігації; методи теорії ймовірностей та математичної статистики – для обробки стохастичних даних та оцінки невизначеності середовища; принципи об'єктно-орієнтованого програмування та імітаційного моделювання – для програмної реалізації та експериментальної апробації системи керування.

Наукова новизна одержаних результатів:

1. Удосконалено модель соціально-адаптивної навігації автономного мобільного робота на основі формалізації марковського процесу прийняття рішень, яка, на відміну від існуючих, завдяки інтеграції кінематичних обмежень робота з параметрами соціальної взаємодії, дозволяє системі керування ідентифікувати та класифікувати потенційні проксемічні конфлікти в режимі реального часу й підвищити безпеку руху в динамічному соціальному середовищі.

2. Вперше розроблено метод навчання навігаційної політики на основі глибинного навчання з підкріпленням, який за рахунок застосування комбінованої стратегії Curriculum Learning із механізмом автоматизованого переходу між етапами складності дозволяє вирішити проблему розрідженої винагороди та забезпечити прискорення збіжності нейромережевої моделі в умовах соціальної навігації.

3. Вперше розроблено метод адаптивного формування винагороди на основі прогнозу невизначеності динамічного середовища, який за рахунок

впровадження механізму динамічного зважування компонентів функції винагороди дозволяє оптимізувати стратегію поведінки автономного мобільного робота залежно від рівня невизначеності середовища.

Практичне значення дисертаційного дослідження. Удосконалена модель соціально-адаптивної навігації автономного мобільного робота, яка завдяки інтеграції кількісних параметрів проксеміки трансформує неявні правила поведінки людей у алгоритмічні обмеження, дозволяє генерувати передбачувані безконфліктні траєкторії руху в соціальному середовищі. Розроблений метод навчання навігаційної політики на основі глибинного навчання з підкріпленням та Curriculum Learning з автоматизованим механізмом переходу між етапами складності, який шляхом послідовної декомпозиції навігаційних завдань забезпечує стабільну збіжність політики керування та зростання частоти досягнення автономним мобільним роботом заданої цілі із одночасним зменшенням кількості фізичних зіткнень. Розроблений метод адаптивного формування винагороди на основі прогнозу невизначеності динамічного середовища, який з урахуванням поточного рівня невизначеності автоматично коригує пріоритети між цільовою ефективністю та соціальною безпекою, надає змогу програмному забезпеченню сервісних мобільних систем генерувати безпечні стратегії руху в стохастичних умовах.

Особистий внесок здобувача. Усі результати, подані до захисту, розроблені автором особисто. У наукових публікаціях, підготовлених у співавторстві, здобувачеві належать такі результати: проведено комплексний аналіз сучасних методів навігації мобільних роботів у закритих приміщеннях, класифіковано підходи до планування траєкторії та визначено обмеження класичних алгоритмів у динамічних середовищах [29], здійснено порівняльний аналіз архітектур глибинного навчання з підкріпленням для задач планування шляху, обґрунтовано вибір алгоритмів для оптимізації руху мобільних роботів та визначено ключові метрики ефективності навчальних стратегій [74], розроблено модель соціально-адаптивної навігації [104], запропоновано метод навчання на основі стратегії

Curriculum Learning [110], розроблено структуру навігаційної системи автономного мобільного робота [129].

Апробація результатів дисертації. Ключові положення та практичні результати дисертаційної роботи пройшли апробацію шляхом представлення на науково-практичних конференціях:

- Всеукраїнська науково-технічна конференція «Технологічні горизонти: дослідження та застосування інформаційних технологій для технологічного прогресу України та світу» (28 листопада 2023 р., м. Київ);
- Всеукраїнська науково-практична конференція «Цифрова гуманістика: Інформаційні технології та інформаційне моделювання на сучасному етапі розвитку суспільства» (4-5 червня 2024 р., м. Кропивницький);
- Наукова конференція молодих вчених «Інновації» (19 вересня 2024 р., м. Київ);
- II Всеукраїнська науково-технічна конференція «Технологічні горизонти: дослідження та застосування інформаційних технологій для технологічного прогресу України і світу» (18 листопада 2024 р., м. Київ);
- V Всеукраїнська науково-практична конференція «Проблеми комп'ютерної інженерії» (03 грудня 2024 р., м. Київ);
- II Міжнародна конференція «Сучасні аспекти діджиталізації та інформатизації в програмній та комп'ютерній інженерії» (19-21 грудня 2024 р., м. Київ);
- Міжнародна наукова конференція «Cybersecurity Providing in Information and Telecommunication Systems» (28 лютого 2025 р., м. Київ);
- II Всеукраїнська науково-практична конференція «Цифрова гуманістика: Інформаційні технології та інформаційне моделювання на сучасному етапі розвитку суспільства» (22-23 травня 2025 р., м. Кропивницький);
- Всеукраїнська науково-технічна конференція «Виклики та рішення в програмній інженерії» (26 листопада 2025 р., м. Київ);
- VI Всеукраїнська науково-практична конференція «Проблеми комп'ютерної інженерії» (03 грудня 2025 р., м. Київ);

- III Міжнародна конференція «Сучасні аспекти діджиталізації та інформатизації в програмній та комп'ютерній інженерії» (4-6 грудня 2025 р., м. Київ).

Структура та обсяг дисертації. Дисертаційна робота складається з анотації, змісту, вступу, чотирьох розділів, загальних висновків, списку використаних джерел та додатків. Робота містить 19 рисунків, 9 таблиць та 8 сторінок додатків. Список використаних джерел налічує 131 найменування. Загальний обсяг дисертації становить 171 сторінка, з них 131 сторінка основного тексту.

РОЗДІЛ 1

СУЧАСНИЙ СТАН ДОСЛІДЖЕНЬ НАВІГАЦІЇ АВТОНОМНИХ МОБІЛЬНИХ РОБОТІВ У ДИНАМІЧНОМУ СЕРЕДОВИЩІ

1.1. Особливості функціонування автономних мобільних роботів у динамічному середовищі

Сучасний етап розвитку робототехніки характеризується стрімкою інтеграцією роботизованих систем у повсякденне життя людини та виробничі процеси в контексті Індустрії 4.0 [1]. Мобільні роботи (МР) еволюціонували від механізмів, які рухаються фіксованими траєкторіями, до автономних агентів, здатних функціонувати в неструктурованих та динамічних середовищах. Сфери їх застосування охоплюють логістику, медицину, сервісне обслуговування, патрулювання та допомогу людям з обмеженими можливостями.

Аналіз фахових джерел свідчить про відсутність уніфікованого термінологічного апарату щодо поняття «робот». Так, згідно з Міжнародною організацією зі стандартизації (ISO), робот визначається як програмований приводний механізм, який володіє певним рівнем автономності та призначений для здійснення просторового переміщення, позиціювання або маніпулювання об'єктами [2].

Американський інститут робототехніки (RIA) трактує робота як перепрограмований багатофункціональний маніпулятор, спроектований для транспортування матеріалів, деталей чи інструментарію за варіативними траєкторіями задля виконання спектра виробничих завдань [3].

Більш антропоморфний підхід пропонує Японська асоціація робототехніки (JARA), яка розглядає робота як механічну систему з гнучкою моторикою, подібною до рухів живих організмів. Згідно з цим визначенням, ключовою ознакою робота є поєднання рухових функцій з інтелектуальними здібностями, такими як розпізнавання образів, адаптація до середовища та здатність до навчання [4].

Автономний мобільний робот (АМР) визначають як систему, яка здатна самостійно виконувати поставлені завдання у середовищі без явного зовнішнього керування [5]. Ключові функціональні можливості АМР базуються на безперервному циклі сенсорного моніторингу простору, інтелектуальній інтерпретації масивів даних, проактивному плануванні поведінки та фізичній взаємодії із середовищем.

Передумовою реалізації зазначеної просторової взаємодії та успішного виконання завдань є безпечне переміщення АМР у середовищі. У цьому контексті навігація автономного мобільного робота розглядається як процес просторової локалізації, формування карти середовища, планування та керування рухом для досягнення заданої цілі з урахуванням наявних перешкод і динамічних змін у середовищі [6].

Функціональна схема процесу навігації АМР є інтеграцією взаємопов'язаних підсистем: сприйняття, локалізації, планування та переміщення [7]. Схему процесу навігації АМР представлено на рис. 1. 1.

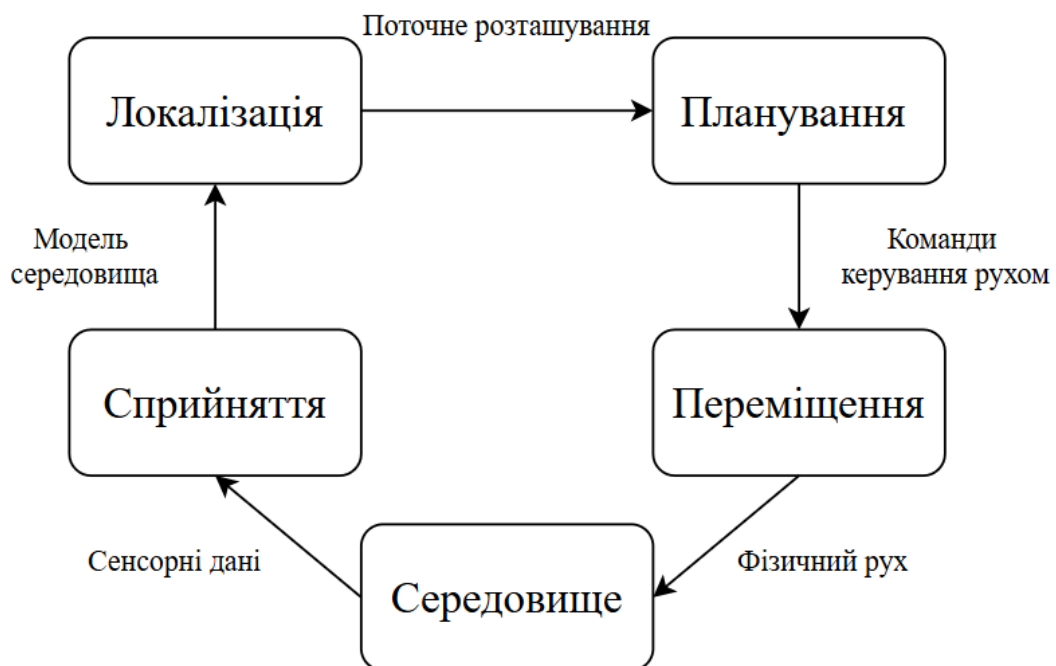


Рис. 1. 1. Функціональна схема процесу навігації АМР

Архітектура та конструктивні особливості АМР орієнтовані на автономну реалізацію навігаційних завдань у різних середовищах експлуатації.

Класифікація середовищ функціонування АМР здійснюється за критеріями апріорної інформації та динамічності об'єктів [8]:

1. Відоме статичне середовище передбачає наявність заздалегідь сформованої карти простору за умови повної відсутності змін під час руху.
2. Відоме динамічне середовище характеризується наявністю карти, проте допускає появу рухомих об'єктів або зміну конфігурації перешкод.
3. Невідоме статичне середовище визначається як простір без апріорної інформації, яке залишається незмінним протягом усього циклу навігації.
4. Невідоме динамічне середовище функціонує в умовах повної відсутності початкових даних та постійних динамічних змін у реальному часі.

Підсистема *сприйняття* є ключовою ланкою навігації АМР, яка забезпечує збір та попередню обробку сенсорних даних для побудови моделі навколишнього середовища. Функціонал цієї підсистеми охоплює спектр завдань від виявлення перешкод до класифікації місцевості та забезпечення людино-машинної взаємодії.

Апаратна реалізація сприйняття базується на комплексі сенсорів, класифікацію яких здійснюють за такими критеріями [9]:

1. За об'єктом вимірювання:
 - пропріоцептивні датчики (гіроскопи, одометри, акселерометри), які здійснюють моніторинг внутрішнього стану системи (кінематичних та динамічних параметрів) [10];
 - екстероцептивні датчики (лідари, RGB/RGB-D камери), які фіксують параметри зовнішнього середовища для виділення ознак сцени [11].
2. За принципом взаємодії із середовищем:
 - пасивні сенсори (мікрофони, термометри) функціонують за принципом одностороннього сприйняття сигналів із зовнішнього середовища [12];
 - активні сенсори (ультразвукові сонари, лазерні далекоміри) працюють за принципом «випромінювання–відбиття».

Вибір певного типу датчиків обумовлюється не лише їхньою точністю, а й стійкістю до зовнішніх впливів. Зокрема, застосування активних сенсорів пов'язане

з ризиком інтерференції та виникненням перехресних завад, здатних спотворювати цільовий сигнал і знижувати надійність системи [9].

Для мінімізації невизначеності сприйняття в автономних системах застосовують імовірнісні методи інтеграції сенсорних даних, які дозволяють синтезувати інформацію з різних джерел для формування достовірної моделі середовища. Завдяки їх застосуванню досягається висока точність реконструкції стану системи та гарантується її стабільне функціонування за умов часткової втрати сенсорної інформації [13].

Необхідною умовою для вирішення задач локалізації та планування руху є наявність *моделі середовища*. У сучасній робототехніці виділяють три фундаментальні підходи до представлення середовища: метричний, топологічний та семантичний [14].

Метричне представлення базується на відтворенні геометричних властивостей середовища у фіксованій системі координат. Найпоширенішою реалізацією є карти сітки зайнятості, де простір дискретизується на комірки, кожна з яких має імовірнісну оцінку наявності перешкоди. Перевагами даного представлення є висока точність опису перешкод, інтуїтивна зрозумілість для оператора, ефективність для завдань локального планування та точного позиціонування. Серед недоліків слід зазначити чутливість до накопичувальних похибок одометрії та необхідність значних обсягів пам'яті при масштабуванні простору [15].

Топологічне представлення моделює середовище у вигляді графа, в якому вузли відповідають певним визначеним місцям (кімнати, перехрестя), а дуги – можливим шляхам переходу між ними. Такий підхід ігнорує точну геометрію на користь структури зв'язків [16]. Топологічне представлення відзначається компактністю зберігання даних, низьким обчислювальним навантаженням, легкістю модифікації графа при зміні структури середовища та ефективністю при плануванні маршрутів на великих дистанціях. Проте даному підходу властива проблема «перцептивного аліасингу» – схожі місця можуть сприйматися як

однакові. Також недоліком є необхідність попереднього фізичного дослідження простору для формування вузлів графа.

Семантичне представлення є вищим рівнем абстракції, який додає до просторових даних змістове навантаження. Семантична карта оперує концептуальними сутностями (об'єкти, функціональні зони, події) та зв'язками між ними. Наприклад, робот не просто бачить перешкоду, а ідентифікує її як «стіл» або «людину» [17]. Перевагою даного представлення є забезпечення контекстного розуміння середовища. Такий підхід гарантує стійкість роботи навігаційних алгоритмів в умовах інформаційної неповноти та стохастичності середовища. Недоліком семантичного представлення є висока залежність від якості алгоритмів розпізнавання образів та значна потреба в обчислювальних ресурсах для семантичної сегментації в реальному часі [18].

Вибір варіанту представлення середовища залежить від специфіки завдань, які виконує автономний мобільний робот. Топологічні моделі є оптимальними для глобальної навігації у великих приміщеннях, оскільки вони імітують людський спосіб орієнтування. Метричні карти залишаються стандартом для точного маневрування та уникнення зіткнень. Водночас, семантичне представлення наближає сприйняття АМР до когнітивних моделей людини.

На основі обраної моделі середовища реалізується процес *локалізації* – визначення розташування робота в робочому просторі. Оскільки глобальні системи позиціонування (GPS) часто недоступні або недостатньо точні в приміщеннях, в таких середовищах застосовують методи відносної локалізації. Проблема невизначеності сенсорних даних при цьому вирішується методами імовірнісної локалізації, такими як фільтр Калмана [19], марковська локалізація [20] або методи Монте-Карло [21].

Процес локалізації характеризується глибоким рекурсивним зв'язком: оцінка позиції робота неможлива без апіорної карти, тоді як побудова достовірної моделі середовища вимагає точних даних про траєкторію руху. Ця класична проблема робототехніки вирішується в рамках парадигми SLAM (Simultaneous Localization and Mapping). SLAM здійснює одночасну оцінку стану робота та

реконструкцію карти [22]. Архітектурна основа сучасних рішень SLAM базується на трьох фундаментальних підходах: імовірнісному EKF-SLAM (на основі розширеного фільтра Калмана), FastSLAM (з використанням фільтра частинок) та методах оптимізації на графах (Graph-based SLAM) [23].

Підсистема *планування* відповідає за формування стратегії переміщення у середовищі на основі доступної інформації про його стан, модель середовища та задану ціль. У межах цієї підсистеми функціонує когнітивний рівень, на якому здійснюється високорівнева інтерпретація семантичної та просторової інформації, визначення цілей навігації та оцінка можливості їх досягнення за наявних умов. На основі отриманих даних модуль планування руху виконує обчислення шляху та траєкторії руху з урахуванням кінематичних і динамічних характеристик робота.

Процес планування руху поділяють на два рівні: планування шляху та планування траєкторії. Планування шляху передбачає пошук оптимального геометричного шляху без урахування часових параметрів руху, зокрема швидкості та прискорення. Під час планування траєкторії здійснюється розрахунок сигналів керування у часі $u(t)$ із врахуванням динаміки та кінематики робота. При цьому цільовою функцією є не лише безпосереднє досягнення заданої позиції, але й оптимізація характеристик руху, наприклад мінімізація часу переміщення, мінімізація енергозатрат та інші.

Уникнення перешкод є реактивною складовою підсистеми планування та реалізується на основі поточних сенсорних даних, отриманих підсистемою сприйняття. Алгоритми уникнення перешкод поділяють на методи, що базуються на карті (map-based) та методи, які не використовують карту (mapless-based).

Методи із використанням карти оперують геометричними моделями середовища та розглядають задачу уникнення перешкод як пошук вільного простору на попередньо сформованій карті. Ключовою характеристикою таких методів є залежність від точності локалізації робота на карті. До методів із використанням карти належать методи на основі сітки зайнятості (Occupancy Grid Approaches) [24], перетин просторово-часових об'ємів (Space-Time Volume Intersection), інтерференція охоплюваних об'ємів (Swept Volume Interference) [25]

Функціонування методів, які не використовують карту, ґрунтується на безпосередній обробці поточних сенсорних даних. До цього класу алгоритмів належать Vector Field Histogram (VFH) [26], Dynamic Window Approach (DWA) [27] та Curvature-Velocity Method (CVM) [28].

Таким чином, підсистема планування виконує координуючу функцію в архітектурі навігації, забезпечуючи узгодженість поведінки робота із поставленими завданнями та поточним станом середовища. Результатом її роботи є формування команд керування руху, які передаються до підсистеми переміщення для безпосередньої реалізації руху.

Конфігурація підсистеми *переміщення* визначається як технічними вимогами до мобільної платформи (маневреність, керованість, енергоефективність, стійкість), так і фізичними властивостями середовища експлуатації (наземне, водне, повітряне). Узагальнену класифікацію мобільних роботів [29] за принципом переміщення у відповідних середовищах представлено на рис. 1.2.

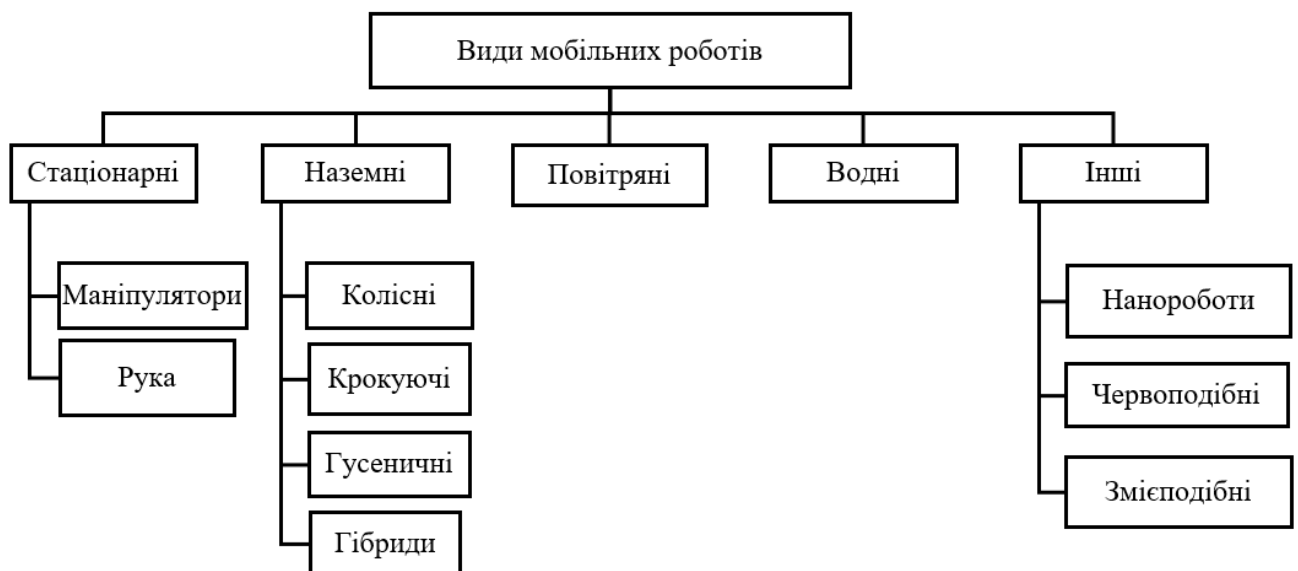


Рис. 1.2. Класифікація мобільних роботів за способом переміщення

Основним викликом для навігаційних систем АМР є перехід від статичних до динамічних середовищ, в яких присутність людей створює високий рівень невизначеності. На відміну від статичних перешкод, люди є активними агентами із стохастичною та важкопрогнозованою поведінкою. У таких умовах класичне

розуміння навігації як процесу переміщення без зіткнень з стартової точки в цільову точку стає недостатнім.

Специфіка функціонування роботів у середовищі з присутністю людей вимагає врахування таких факторів, як стохастичність руху людей та соціальна прийнятність поведінки робота.

Соціальна прийнятність поведінки робота передбачає, що робот повинен діяти відповідно до соціальних норм, його поведінка повинна бути інтуїтивно зрозумілою та комфортною для оточуючих. Це зумовлює необхідність використання систем на основі розуміння людських намірів, імовірнісного прогнозування та адаптивного прийняття рішень в реальному часі.

1.2. Аналіз методів планування руху мобільних роботів у динамічному середовищі

Проблема навігації автономних мобільних роботів є комплексною науково-технічною задачею, яка вимагає ієрархічного підходу до організації руху. У сучасній робототехніці прийнято розмежовувати процес планування на два рівні: глобальне планування та локальне планування.

Глобальне планування базується на наявності апріорної інформації про середовище (відома карта) і полягає у пошуку оптимального маршруту від початкової точки до цільової відповідно до заданих критеріїв ефективності, таких як мінімальний час, найменша енергоємність або найкоротша відстань. На цьому рівні робот оперує повною моделлю статичного середовища, розраховуючи траєкторію руху.

Специфіка локального планування, порівняно з глобальним, полягає у здатності здійснювати динамічне коригування маршруту в реальному часі. Основним завданням локального планування є коригування траєкторії для уникнення зіткнень з перешкодами, інформація про які була відсутня на етапі глобального планування. Локальні методи ефективні для АМР у середовищах, які

змінюються у часі, оскільки вони вимагають менше попередніх знань і базуються на реактивній поведінці.

Класифікацію методів планування шляху АМР представлено на рис. 1. 3.

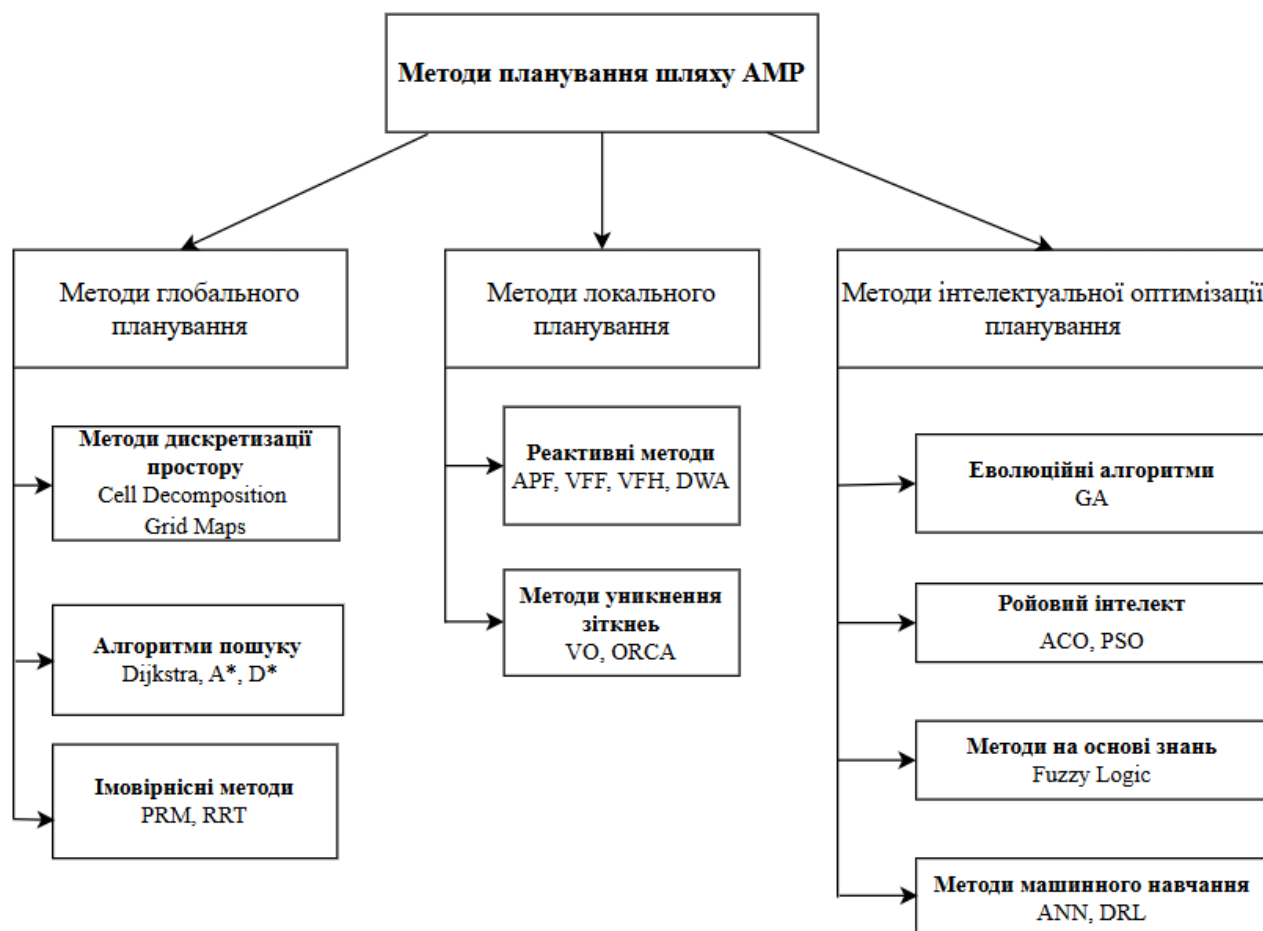


Рис. 1.3. Класифікація методів планування шляху АМР

1.2.1. Методи глобального планування шляху

Процес глобального планування зазвичай включає два ключові етапи: побудову моделі середовища або представлення вільного простору та пошук оптимального шляху в отриманій структурі. Представлення середовища може здійснюватися різними способами, зокрема шляхом дискретизації простору або побудови топологічних структур, які відображають доступні області переміщення робота.

До групи класичних методів дискретизації простору належать метод клітинної декомпозиції (Cell Decomposition) [30] та методи побудови дорожніх карт (Roadmap methods) [31].

Метод клітинної декомпозиції [32] полягає у розбитті конфігураційного простору C (C -space) на скінченну кількість областей c_i , які не перетинаються і належать вільному простору:

$$C_{free} = \bigcup_{i=1}^n c_i, \quad (1.1)$$

де $c_i \cap c_j = \emptyset$ і $c_i \subset C_{free}$.

Залежно від способу формування меж та рівня деталізації вільного простору, методи клітинної декомпозиції поділяють на три основні категорії: точну, наближену та адаптивну.

Точна декомпозиція здійснює поділ середовища на основі геометричного розташування перешкод. Межі клітинок визначаються вершинами та ребрами багатокутників. У результаті формується граф суміжності клітин, в якому кожна клітинка відповідає вузлу графа, а ребра відображають можливість переходу між сусідніми областями. Така модель забезпечує компактне та точне представлення простору. Однак практичне застосування точної декомпозиції малоефективне в динамічних сценаріях через складність перерахунку геометричних меж при зміні положення об'єктів [33].

Наближена декомпозиція використовує сітки зайнятості (Occupancy Grids) з фіксованою роздільною здатністю [34]. У багатьох сучасних реалізаціях кожна комірка такої сітки зберігає імовірнісну оцінку наявності перешкоди, дозволяючи роботу поступово уточнювати карту середовища на основі сенсорних даних. Такий підхід є одним із найпоширеніших методів інтеграції сенсорної інформації. Проте він має суттєвий недолік – вибір зовеликої клітинки може призвести до хибної класифікації вузьких проходів як заблокованих.

Адаптивна декомпозиція базується на рекурсивному поділі простору, наприклад, за допомогою структури квадродерев. Простір розділяється на окремі квадранти лише в тому випадку, якщо клітинка одночасно містить фрагменти вільного простору і перешкоди. Такий підхід дозволяє досягти високої деталізації в критичних зонах, наприклад поблизу кутів або у вузьких проходів. Водночас

зберігається низький рівень дискретизації для великих відкритих областей. Забезпечується компроміс між точністю представлення середовища та обсягом використаної оперативної пам'яті комп'ютерної системи робота [35].

Таким чином, вибір конкретної стратегії декомпозиції визначає не лише точність планування шляху, а й здатність робота адаптуватися до змін середовища. У класичних геометричних підходах до планування шляху часто використовують діаграму Вороного, яка максимізує відстані до найближчих перешкод [36]. У цьому випадку шлях будується вздовж множини точок, рівновіддалених від двох найближчих об'єктів:

$$V(O) = \{q \in C_{free} | \|q - P_i\| = \|q - P_j\| \leq \|q - P_k\|, \forall k \neq i, j\}, \quad (1.2)$$

де P_i – найближча точка на поверхні перешкоди O_i .

Такий підхід дозволяє мінімізувати ризик фізичного зіткнення з об'єктами середовища. Проте у соціальному контексті робот може ігнорувати комфортну дистанцію взаємодії з людьми, орієнтуючись лише на геометричний центр проходу.

Метод PRM (Probabilistic Roadmaps) [37] ефективний для багатовимірних конфігураційних просторах. Цей метод складається з двох основних етапів: побудови дорожньої карти та пошуку шляху. На етапі побудови графа випадково генерується множина конфігурацій Q_{rand} , які перевіряються на належність до вільного простору:

$$Q_{valid} = \{q \in Q_{rand} | q \in C_{free}\}. \quad (1.3)$$

З'єднання вузлів q_i та q_j відбувається лише за умови, що локальний планувальник $L(q_i, q_j)$ підтверджує відсутність перешкод на відрізку між ними. Метод PRM ефективний у статичних середовищах, в яких дорожня карта може бути побудована заздалегідь. Проте в динамічному середовищі функція перевірки зіткнень $Coll(L(q_i, q_j))$ швидко втрачає актуальність через переміщення об'єктів, зокрема людей.

Недоліком PRM є залежність від попередньо побудованого графа. Динамічне переміщення людей може стати причиною раптового блокування вузлів і ребер

дорожньої карти, які попередньо були ідентифіковані як вільні. Така мінливість середовища потребує безперервної перебудови графа, знижує швидкість методу та обмежує його застосування у високодинамічних соціальних сценаріях.

Після формування графа або сітки застосовуються алгоритми пошуку для знаходження найкоротшої траєкторії. Основними є алгоритми Дейкстри [38], A* (A-star) [39] та D* (Dynamic A*) [40].

Алгоритм Дейкстри виконує детермінований пошук найкоротшого шляху у графі та фактично мінімізує накопичену вартість маршруту. Натомість алгоритм A* використовує евристичну оцінку для прискорення пошуку [41]. У цьому випадку вибір наступного вузла n здійснюється на основі цільової функції:

$$f(n) = g(n) + h(n), \quad (1.4)$$

де $g(n)$ – накопичена вартість шляху від початкового вузла до вузла n ;

$h(n)$ – евристична оцінка вартості шляху від поточного вузла до цільової точки.

У більшості класичних реалізацій функція $g(n)$ визначається геометричною довжиною траєкторії. Однак у соціальних задачах така метрика не враховує додаткову «соціальну вартість» перебування робота поблизу людини.

Зазначені методи демонструють високу ефективність у знаходженні найкоротшої траєкторії за умови, що карта середовища є повністю відомою та незмінною. Проте в умовах соціального середовища поведінка людей є стохастичною, а геометрія вільного простору змінюється безперервно. Тому детерміновані алгоритми планування не завжди здатні враховувати динамічні зміни середовища. У таких умовах класичні планувальники змушені часто виконувати перепланування маршруту, що може призводити до нестабільного або переривчастого руху робота.

Модифікації класичного алгоритму пошуку, такі як D* [42], дозволяють проводити локальну корекцію шляху при виявленні нових перешкод без повного перепланування. Проте ці методи оптимізують переважно геометричні параметри (довжину шляху, відсутність зіткнень) і не враховують соціальні норми взаємодії.

У результаті траєкторія, побудована такими алгоритмами, може проходити занадто близько до людини.

Таким чином, класичні методи глобального планування забезпечують геометричну безпеку та математичну оптимальність траєкторій. Водночас їх адаптивність обмежена, оскільки вони не враховують динамічну природу соціального середовища та особливості взаємодії з людьми.

1.2.2. Методи локального планування траєкторії

Для навігації в реальному часі та уникнення зіткнень застосовують реактивні методи. На відміну від глобальних планувальників, вони використовують поточну сенсорну інформацію про стан середовища, яка забезпечує високу швидкість реакції на динамічні зміни.

Одним із класичних методів локального планування є метод штучних потенційних полів (Artificial Potential Field, APF) [43]. В основі методу лежить моделювання мобільного робота як матеріальної точки, яка рухається у віртуальному силовому полі [44].

Середовище моделюється як сукупність двох типів сил – сили притягання (Attractive Force), яка генерується ціллю, та сили відштовхування (Repulsive Force), яку створюють перешкоди.

Якщо $q = (x, y)^T$ – вектор поточного розташування робота. Тоді функція штучного потенційного поля $U(q)$ визначається як сума потенціалів притягання та відштовхування

$$U(q) = U_{att}(q) + U_{rep}(q), \quad (1.5)$$

де $U_{att}(q)$ – потенціал притягання до цілі;

$U_{rep}(q)$ – потенціал відштовхування від перешкод.

Сила $F(q)$, яка діє на робота, визначається як негативний градієнт потенціалу

$$F(q) = -\nabla U(q). \quad (1.6)$$

Рівнодійна сила визначає напрямок руху робота.

Метод APF характеризується низькою обчислювальною складністю та високою швидкістю, що робить його придатним для задач локального планування у реальному часі.

Попри простоту реалізації, метод має суттєві обмеження для соціальної навігації. Основною проблемою є виникнення локальних мінімумів, коли результуюча сила рівна нулю і призводить до зупинки робота [45]. Крім того, у динамічному середовищі APF також може демонструвати нестабільну поведінку. Наприклад, при взаємодії з рухомими людьми робот може коливатися між напрямками обходу перешкоди, що призводить до осциляцій траєкторії.

Еволюційним розвитком методу потенційних полів став метод віртуального силового поля (Virtual Force Field, VFF), запропонованого Й. Боренштейном і Й. Кореном [46]. VFF інтегрує концепцію потенційних полів із методом ймовірнісних сіток зайнятості.

Основна ідея методу полягає в ітеративному оновленні ймовірнісної карти простору шляхом обробки сигналів, які надходять від ультразвукових далекомірів. Сформована у такий спосіб локальна карта слугує базисом для розрахунку потенційного силового поля, вплив якого діє на платформу робота.

Робочий простір представляється у вигляді двовимірної метричної сітки. Кожній комірці сітки $C(i, j)$ присвоюється значення впевненості щодо наявності перешкоди в діапазоні $0 \leq C(i, j) \leq C_{max}$, де 0 відповідає вільному простору, а C_{max} – гарантованій наявності перешкоди в заданій комірці.

Для розрахунку керуючого сигналу навколо робота формується активне вікно, яке рухається разом із роботом. Кожна комірка всередині цього вікна створює віртуальний вектор відштовхування

$$F_{i,j} = \frac{F_{cr} \cdot C(i, j)}{d^2(i, j)} \cdot n_{i,j}, \quad (1.7)$$

де F_{cr} – коефіцієнт сили відштовхування; $C(i, j)$ – значення впевненості комірки (рівень небезпеки); $d(i, j)$ – відстань від робота до комірки (i, j) ; $n_{i,j}$ – одиничний вектор, спрямований від комірки до робота.

Результуюча сила відштовхування F_r є сумою сил від усіх комірок активного вікна

$$F_r = \sum_{i,j \in W} F_{i,j}. \quad (1.8)$$

Сила притягання F_t створюється цільовою точкою. Вона має постійну величину (магнітуду), коли робот знаходиться далеко від цілі, і зменшується пропорційно відстані при наближенні до неї:

$$F_t = K_t(x_t - x_r), \quad (1.9)$$

де x_t – положення цілі; x_r – положення робота.

Вектор руху R обчислюється як векторна сума сили притягання до цілі та сили відштовхування від перешкод:

$$R = F_t + F_r. \quad (1.10)$$

Напрямок вектора R визначає кут повороту робота θ_{ref} , а його величина може використовуватися для регулювання швидкості.

Незважаючи на простоту реалізації, метод VFF демонструє нестабільність руху робота у вузьких проходах та може призводити до виникнення локальних мінімумів [47].

Для усунення недоліків VFF Й. Боренштейн запропонував метод гістограми векторного поля (Vector Field Histogram, VFH) [48].

Метод VFH використовує трирівневу репрезентацію даних:

- двовимірну декартову гістограму;
- одновимірну полярну гістограму;
- обчислення командних сигналів динамічного регулювання.

Ключовою інновацією методу стало перетворення локальної карти перешкод у полярну гістограму H навколо робота. Вміст кожної активної клітинки (i,j) трактується як вектор перешкоди, що характеризується напрямком $\beta_{i,j}$ від клітинки до центру робота, магнітудою $m_{i,j}$, яка залежить від значення впевненості клітинки $c_{i,j}^*$ та квадрата відстані $d_{i,j}$

$$m_{i,j} = (c_{i,j}^*)^2 \cdot (a - b \cdot d_{i,j}), \quad (1.11)$$

де $a - b \cdot d_{i,j} \geq 0$ і a, b – константи, підібрані так, щоб магнітуда була максимальною біля робота і дорівнювала нулю на межі активного вікна.

Кожна клітинка (i, j) додається до відповідного сектора k полярної гістограми. Сектор визначається як

$$k = INT\left(\frac{\beta_{ij}}{\alpha}\right), \quad (1.12)$$

де α – кутова роздільна здатність гістограми.

Значення полярної щільності перешкод h_k для сектора k є сумою магнитуд усіх клітинок, які потрапляють у цей сектор:

$$h_k = \sum m_{i,j}. \quad (1.13)$$

Для усунення шуму дискретизації застосовується згладжування гістограми H за формулою ковзного середнього:

$$h'_k = \frac{1}{2l+1} \sum_{i=-l}^l h_{k+i}. \quad (1.14)$$

На основі згладженої гістограми визначаються вільні сектори руху.

На відміну від VFF, де швидкість часто була постійною, VFH динамічно знижує швидкість при наближенні до перешкод. Швидкість V зменшується пропорційно щільності перешкод у напрямку руху h'_c

$$V' = V_{max} \cdot \left(1 - \frac{h'_c}{h_m}\right), \quad (1.15)$$

де h_m – максимальна щільність перешкод, при якій робот зупиняється. Додатково швидкість знижується при виконанні різких поворотів.

Попри спроможність методів VFF та VFH забезпечувати рух у середовищах із статичними та динамічними об'єктами, ці методи демонструють низку специфічних недоліків, які обмежують їхню придатність для задач у середовищі з людьми.

У соціальному середовищі застосування методу віртуального силового поля супроводжується проблемою нестабільності руху, зумовленою дискретною природою ймовірнісної сітки. У динамічному натовпі люди постійно змінюють своє положення, тому оновлення даних у комірках сітки створює значний інформаційний шум і призводить до різких та частих змін векторів рівнодійної сили. Це спричиняє осциляцію траєкторії мобільного робота.

Метод гістограми векторного поля, попри свою ефективність у швидкісному маневруванні, демонструє конфлікт між алгоритмічною оптимізацією вільного простору та соціальними нормами, оскільки VFH схильний спрямовувати робота у будь-які геометрично доступні проходи, ігноруючи їх соціальний контекст. Це може призвести до критичного наближення до людини, якщо алгоритм визначить таку траєкторію як найбільш ефективну на гістограмі. Крім того, здатність VFH підтримувати високу швидкість у вузьких проходах, яка вважається технічною перевагою методу, у соціальному середовищі перетворюється на ризик, оскільки швидкий рух у безпосередній близькості до людини викликає відчуття загрози.

Суттєвим недоліком методів на основі потенційних полів та гістограм є нехтування динамічними характеристиками самого робота. Оскільки ці алгоритми розраховують лише геометрично оптимальний вектор руху без урахування інерції та обмежень приводу, реальна траєкторія часто відхиляється від розрахованої і створює загрозу зіткнення у обмеженому просторі.

На противагу зазначеним підходам, метод динамічного вікна (Dynamic Window Approach, DWA) забезпечує генерацію плавних траєкторій. Інтегруючи кінематичні обмеження робота безпосередньо в процес обчислення швидкостей, алгоритм усуває проблему розривності керування.

Метод DWA, запропонований Д. Фоксом, В. Бургардом та С. Труном [27], є підходом до реактивного уникнення зіткнень і функціонує безпосередньо у просторі швидкостей (v, w) робота. На відміну від інших методів, DWA явно враховує інерційні характеристики та кінематичні обмеження платформи [49], апроксимуючи траєкторії руху дугами кіл.

Алгоритм базується на циклічному пошуку оптимальної команди у просторі швидкостей V_r , який формується як перетин трьох множин обмежень: технічних можливостей двигунів, динамічних характеристик прискорення та умов безпеки гальмування.

Множина всіх можливих швидкостей V_s визначається технічними характеристиками приводів робота і обмежує максимальну та мінімальну лінійну v та кутову ω швидкості

$$V_s = \{(v, \omega) \mid v \in [v_{min}, v_{max}] \wedge \omega \in [\omega_{min}, \omega_{max}]\}. \quad (1.16)$$

Множина динамічного вікна V_d включає лише ті пари швидкостей, які можуть бути досягнуті протягом наступного короткого часового інтервалу t з поточних значень (v_a, ω_a) при максимальних прискореннях лінійного руху a та кутового руху ε .

$$V_d = \{(v, \omega) \mid v \in [v_a - a_{br} \cdot t, v_a + a_{acc} \cdot t] \wedge \omega \in [\omega_a - \varepsilon_{br} \cdot t, \omega_a + \varepsilon_{acc} \cdot t]\}, \quad (1.17)$$

де a_{acc} , a_{br} – максимальні лінійні прискорення розгону та гальмування, а ε_{acc} , ε_{br} – відповідні максимальні кутові прискорення.

До множини допустимих швидкостей V_a входять виключно ті пари (v, ω) , за яких робот здатен повністю зупинитися до зіткнення з найближчою перешкодою на розрахованій криволінійній траєкторії. Умова безпеки формалізується нерівністю:

$$V_a = \{(v, \omega) \mid v \leq \sqrt{2dist(v, \omega) \cdot a_{br}} \wedge \omega \leq \sqrt{2dist(v, \omega) \cdot \varepsilon_{br}}\}, \quad (1.18)$$

де $dist(v, \omega)$ – відстань до найближчої перешкоди вздовж дуги траєкторії.

Остаточний простір пошуку V_r , у якому здійснюється вибір керування, є перетином зазначених множин:

$$V_r = V_s \cap V_a \cap V_d. \quad (1.19)$$

Вибір оптимальної пари швидкостей (v, ω) із V_r здійснюється шляхом максимізації функції $G(v, \omega)$, яка враховує навігаційні цілі та обмеження середовища.

Цільова функція має вигляд

$$G(v, \omega) = \sigma(\alpha \cdot \text{angl}(v, \omega) + \beta \cdot \text{dist}(v, \omega) + \gamma \cdot \text{vel}(v, \omega)), \quad (1.20)$$

де σ – функція згладжування; $\text{angl}(v, w)$ – метрика орієнтації, яка визначається як $180 - |\theta_{\text{target}} - \theta_{\text{predicted}}|$; $\text{dist}(v, w)$ – значення мінімальної дистанції до об'єктів середовища вздовж дуги траєкторії; $\text{vel}(v, w)$ стимулює вибір максимально можливої безпечної швидкості.

Вагові коефіцієнти α , β , γ дозволяють налаштовувати поведінку робота, балансуючи між агресивним рухом до цілі та обережним маневруванням.

Метод DWA базується на агенто-центричному підході. У межах цієї парадигми всі навколишні об'єкти, включаючи людей, сприймаються як пасивні перешкоди, які не реагують на присутність робота. Однак у реальному соціальному середовищі процес уникнення зіткнень є інтерактивним та взаємним: людина також коригує власну траєкторію, передбачаючи маневр технічного засобу.

Для подолання обмежень суто реактивного підходу та врахування кооперативної природи руху в натовпі було розроблено групу методів на основі концепції взаємної відповідальності агентів.

Провідне місце серед них посідає метод оптимального взаємного уникнення зіткнень (Optimal Reciprocal Collision Avoidance, ORCA) [50]. Даний алгоритм розширює концепцію перешкод у просторі швидкостей (Velocity Obstacles, VO) та базується на принципі взаємної відповідальності агентів за запобігання зіткнення.

Математично метод ORCA для двох агентів А та В визначає множину допустимих швидкостей через побудову півплощини у просторі відносних швидкостей. Якщо $VO_{A|B}^{\tau}$ – це множина швидкостей агента А, які призведуть до зіткнення з агентом В протягом часу τ , то для будь-якої поточної відносної швидкості $v_A - v_B$, алгоритм знаходить мінімальну зміну швидкості u :

$$u = (\text{argmin}_{v \in \partial VO} \|v - (v_A - v_B)\|) - (v_A - v_B), \quad (1.21)$$

де ∂VO – межа простору заборонених швидкостей.

Перевага ORCA полягає у розподілі зусиль між агентами. Кожен агент змінює свою швидкість лише на половину необхідного вектора u

$$v_A^{new} = \{v | (v - (v_A + \frac{1}{2}u)) \cdot n \geq 0\}, \quad (1.22)$$

де n – нормаль до межі VO.

Така взаємність гарантує відсутність осциляцій та стабільність системи при великій кількості агентів. Проте недоліком ORCA є припущення про ідеальну взаємність поведінки, що не завжди відповідає реальній поведінці людей.

Недоліки реактивних методів спонукали дослідників до інтеграції соціальних норм у математичні моделі навігації. Однією з таких концепцій є модель соціальних сил (Social Force Model, SFM), запропонована Д. Хельбінгом та П. Молнаром [51]. Модель соціальних сил у контексті мобільної робототехніки використовується як реактивний механізм формування швидкісних команд з урахуванням соціальної взаємодії [52].

У моделі рух агента описується як зміна його швидкості v_i під дією суми сил:

$$m_i \frac{dv_i}{dt} = F_i^{goal} + \sum_{j \neq i} F_{ij}^{rep} + \sum_w F_{iw}^{obs}, \quad (1.23)$$

де F_i^{goal} – сила руху до цілі, яка визначає бажання агента підтримувати певну швидкість v_i^0 у напрямку e_i за поточної швидкості v_i з часом релаксації τ (чим менше τ , тим швидше робот реагує на зміну напрямку та набирає швидкість):

$$F_i^{goal} = \frac{v_i^0 e_i - v_i}{\tau}, \quad (1.24)$$

F_{ij}^{rep} – соціальна сила відштовхування між агентами i та j , що експоненціально залежить від відстані d_{ij} :

$$F_{ij}^{rep} = A e^{\frac{r_{ij} - d_{ij}}{B}} n_{ij}, \quad (1.25)$$

де A – сила відштовхування, B – радіус дії сили, n_{ij} – вектор напрямку від i до j .

F_{iw}^{obs} – сила відштовхування від статичних перешкод. На відміну від соціальної сили, вона має вищу інтенсивність, щоб гарантувати фізичну цілісність:

$$F_{iw}^{obs} = U_w e^{\left(\frac{r_i - d_{iw}}{B_w}\right)} \cdot n_{iw}, \quad (1.26)$$

де d_{iw} – найкоротша відстань до найближчої точки перешкоди w , n_{iw} – нормаль від поверхні перешкоди до робота, U_w та B_w – параметри інтенсивності, які вищі за соціальні коефіцієнти, щоб робот ніколи не торкався стін.

Модель SFM стала стандартом для симуляції натовпу, але її використання для керування роботами виявило суттєві недоліки: складність налаштування параметрів та високу обчислювальну складність. Коефіцієнти вимагають ручного калібрування для кожного типу середовища. Оскільки параметри, які є ефективними у вільному просторі, можуть призвести до агресивної або неадекватної поведінки у вузьких коридорах.

Модель має високу обчислювальну складність ($O(N^2)$), оскільки для розрахунку руху робота необхідно обчислити сили взаємодії з кожним агентом. Відповідно в реальному часі при великій кількості людей створюються затримки у прийнятті рішень роботом.

Проведений аналіз методів глобального та локального планування дозволяє виділити ряд системних обмежень, які унеможливають їх ефективне застосування для побудови автономних соціально-адаптивних роботів.

По-перше, недоліком є відсутність передбачуваності. Більшість класичних алгоритмів функціонують за реактивним принципом і не здатні прогнозувати довгострокові наміри людини або зміни її траєкторії внаслідок взаємодії з роботом. У реальних сценаріях це призводить до того, що робот реагує на дії людини із запізненням, замість того, щоб діяти на випередження. Цей розрив між вимірюванням та реакцією у динамічному середовищі створює критичні ризики для безпеки.

По-друге, при використанні даних методів не враховується психологічний комфорт людини. Класичні метрики оптимізації фокусуються на мінімізації довжини шляху, часу руху або енерговитрат за умови відсутності фізичного зіткнення з перешкодою. Проте траєкторія, яка є математично оптимальною та безпечною фізично, може бути неприйнятною соціально. Різкі маневри та прискорення руху поблизу людей викликають у них відчуття небезпеки.

Формалізувати таку поведінку через правила («if-then») неможливо через високу варіативність соціальних контекстів.

В таблиці 1.1. проаналізовано ефективність розглянутих алгоритмів планування шляху в контексті соціальної взаємодії.

Таблиця 1. 1

Оцінка ефективності алгоритмів планування шляху в контексті соціальної взаємодії

Метод	Основна концепція	Переваги	Обмеження в соціальному середовищі
A^* , D^*	Пошук найкоротшого шляху на графі	Оптимальність шляху	Нездатність враховувати динаміку натовпу; часте перепланування; неприродні траєкторії
APF	Рух у віртуальному силовому полі	Простота реалізації; висока обчислювальна ефективність	Проблема локальних мінімумів; осциляції поблизу людей
DWA	Оптимізація швидкостей у динамічному вікні	Врахування кінематики робота; уникнення зіткнень	Розглядає людей як статичні перешкоди; ігнорує соціальний комфорт
SFM	Моделювання «соціальних сил» взаємодії	Врахування базових соціальних норм (дистанція)	Зростання обчислювальних витрат при збільшенні кількості агентів
ORCA	Взаємне уникнення зіткнень у просторі швидкостей	Ефективність для великої кількості агентів; уникнення зіткнень	Механічна поведінка; вимога, щоб інші агенти також слідували правилам взаємності (нереалістично для людей)

Узагальнюючи порівняльний аналіз, наведений у таблиці 1.1, можна відзначити системну невідповідність класичних підходів вимогам соціальної навігації. Глобальні планувальники, забезпечуючи геометричну оптимальність шляху, втрачають ефективність через нездатність адаптуватися до стохастичної динаміки натовпу. Реактивні методи гарантують швидкодію, проте досягають безпеки ціною ігнорування соціального комфорту.

Зазначені обмеження актуалізують необхідність пошуку альтернативних підходів, здатних подолати розрив між логікою геометричних алгоритмів та непередбачуваністю людської поведінки.

1.2.3. Інтелектуальні підходи до просторової навігації та моделювання взаємодій

На відміну від класичних алгоритмів планування, методи планування на основі інтелектуальних обчислень не прив'язані до конкретного рівня навігаційної архітектури. Такі методи можуть застосовуватися як для задач глобального планування маршруту, так і для локального планування траєкторії. Зазначені підходи орієнтовані на формування адаптивної навігаційної поведінки АМР [53].

На етапі розвитку інтелектуальної навігації широкого застосування набули методи нечіткої логіки [54], еволюційні алгоритми, зокрема генетичні алгоритми [55], а також підходи на основі ройового інтелекту, до яких належать оптимізація рою частинок (PSO) [56], мурашині алгоритми (ACO) [57] та алгоритми бджолоїної колонії [58].

Системи нечіткої логіки дозволяють формалізувати евристичні правила руху, наближені до людського мислення, проте складність математичного доведення стійкості та валідації таких систем ускладнює їх використання у соціальному середовищі.

Незважаючи на здатність еволюційних та ройових алгоритмів ефективно розв'язувати задачі оптимізації у складних багатовимірних просторах, їх застосування в задачах соціальної навігації мобільних роботів має низку принципових обмежень. Ці обмеження пов'язані як з особливостями стохастичного пошуку, так і з вимогами до роботи навігаційних систем у реальному часі та в умовах безпосередньої взаємодії з людьми.

Головним недоліком еволюційних методів у контексті соціальної навігації є стохастична природа та відсутність гарантій швидкої збіжності в умовах жорстких часових обмежень [59]. Процес навчання або оптимізації таких алгоритмів, як правило, вимагає значної кількості ітерацій і багаторазових оцінок функції

пристосованості, що істотно ускладнює їх використання для планування руху в реальному часі. Крім того, отримана в результаті оптимізації поведінка мобільного робота часто має низький рівень інтерпретованості, який ускладнює формальну верифікацію безпеки та створює додаткові ризики при взаємодії робота з людьми у соціальному середовищі.

Це зумовило зміщення наукового фокусу на методи інтелектуального навчання, які здатні самостійно виділяти складні патерни взаємодії безпосередньо з досвіду або демонстрацій, зокрема методи глибинного навчання з підкріпленням (DRL). Ключовим викликом для сучасних систем є не лише уникнення зіткнень, а й розуміння соціального контексту та намірів пішоходів. У цьому контексті найбільш перспективними є методи CADRL, LSTM-RL, SARL.

На відміну від еволюційних алгоритмів, які потребують великої кількості ітерацій для кожної нової сцени, методи глибинного навчання з підкріпленням, такі як CADRL [60] дозволяють перенести основне обчислювальне навантаження на етап офлайн-навчання.

У контексті соціальної навігації цей підхід пропонує розглядати людей не як статичні перешкоди, а як інтелектуальних агентів, які приймають рішення. Головною перевагою методу для роботи в соціальному просторі є відсутність необхідності в комунікації або попередньому знанні намірів людей, що дозволяє роботу діяти автономно, базуючись лише на спостереженнях за динамікою оточуючих через бортові сенсори.

В основі архітектури методу лежить використання мережі цінності, яка навчається прогнозувати взаємодію робота з людьми. Вхідний вектор мережі містить параметри стану робота та спостережувані характеристики найближчих людей (їхні позиції, швидкості та радіуси особистого простору). Критично важливим аспектом для соціальної навігації є те, що в процесі навчання на мільйонах епізодів система засвоює концепцію неявної взаємності. Це дозволяє роботу не просто уникати зіткнень, а й очікувати, що людина також може незначно скорегувати свій шлях. Такий підхід запобігає неприродній поведінці та зупинкам робота перед пішоходом.

Інтеграція підходу CADRL у навігаційну систему робота надає суттєві переваги.

По-перше, децентралізована природа методу робить його стійким до проблем зі зв'язком та дозволяє легко масштабувати кількість агентів у системі без експоненційного зростання обчислювальної складності.

По-друге, перенесення основних обчислень на етап офлайн-навчання нейронної мережі забезпечує високу швидкодію в реальному часі і дозволяє використовувати метод навіть на обчислювально обмежених вбудованих платформах. Крім того, завдяки навчанню в стохастичному середовищі, метод демонструє вищу робастність до сенсорного шуму порівняно з геометричними алгоритмами, такими як ORCA.

Водночас метод має певні обмеження, які звужують сферу його застосування. Недоліком є використання повнозв'язної нейронної мережі з фіксованим розміром вхідного вектора. Це змушує систему враховувати лише фіксовану кількість найближчих сусідів, ігноруючи інших учасників руху, що може призвести до небезпечних ситуацій у щільному натовпі. Також метод вимагає повної спостережуваності параметрів (позиції та швидкості) сусідніх агентів, що не завжди можливо забезпечити в реальних умовах через перекриття об'єктів сенсорами. Ці обмеження стали поштовхом для подальшого розвитку методів на основі рекурентних мереж та механізмів уваги.

Для подолання проблеми обробки змінної кількості динамічних перешкод та врахування часових залежностей руху було запропоновано інтеграцію рекурентних нейронних мереж (RNN), зокрема архітектури LSTM, у фреймворк навчання з підкріпленням [61].

Цей клас методів, часто згадуваний як LSTM-RL або GA3C-CADRL, використовує здатність LSTM запам'ятовувати послідовності станів. Замість того, щоб реагувати лише на миттєве положення перешкоди, робот кодує історію руху кожного пішохода у прихований вектор стану фіксованої довжини. Це дозволяє ефективно обробляти довільну кількість сусідніх агентів шляхом послідовної обробки їхніх станів, прогнозувати наміри людей на основі їхньої попередньої

траєкторії, вирішувати проблему «застигання» робота у щільному натовпі завдяки кращому розумінню динаміки сцени.

Логічним розвитком методів навігації на базі LSTM стала інтеграція механізмів уваги, реалізована в методі SARL [62]. Основна ідея SARL полягає в тому, що не всі люди в натовпі є однаково важливими для прийняття навігаційного рішення роботом.

Ключовим компонентом методу є спеціалізована архітектура нейронної мережі цінності, яка навчається оцінювати оптимальність дій робота, використовуючи модуль «соціальної уваги». Процес обробки інформації в такій архітектурі починається з кодування станів, де на вхід мережі подається спільний вектор, що включає стан робота (позицію, швидкість, радіус, кінцеву мету) та спостережувані стани всіх n пішоходів.

Перший шар мережі – багатошаровий перцептрон – трансформує кожну пару «робот-людина» у вектор фіксованої довжини, виділяючи базові ознаки взаємодії. Надалі отримані вектори паралельно обробляються двома окремими підмережами: перша (ρ_f) виділяє ознаки парної взаємодії, а друга (φ_α) обчислює «бали уваги» для кожної людини, які відображають відносну важливість конкретної людини для робота в даний момент часу. Фінальний вектор, який описує всіх людей, формується як зважена сума ознак взаємодії всіх сусідів, де вагами виступають розраховані бали уваги після їх нормалізації функцією softmax. Такий механізм дозволяє автоматично агрегувати інформацію про довільну кількість людей у вектор фіксованого розміру.

На основі сформованого вектора контексту та власного стану робота фінальний модуль мережі апроксимує функцію цінності $V^*(J_t)$, яка прогнозує очікуваний час до досягнення цілі та ймовірність зіткнення.

На кожному часовому кроці робот обирає дію у вигляді лінійної та кутової швидкостей, яка максимізує функцію цінності, забезпечуючи найшвидший рух за умови дотримання безпеки. Запропонований підхід дозволяє агенту DRL фокусувати обмежені обчислювальні ресурси на критично важливих об'єктах, автоматично надаючи високу вагу людям, що знаходяться в безпосередній

близькості. Тоді як віддаленим об'єктам присвоюється низька вага. Завдяки цьому SARL формує безпечні та соціально прийнятні траєкторії, імітуючи природну поведінку людини.

Незважаючи на значні переваги у моделюванні соціальної уваги, метод SARL має низку критичних недоліків, які обмежують його безпосереднє застосування в реальних робототехнічних системах. Однією з головних проблем є ігнорування статичної геометрії середовища, оскільки оригінальна архітектура фокусується виключно на взаємодії «робот-людина». Внаслідок цього, у складних приміщеннях з коридорами та меблями робот може успішно уникати людей, але зіштовхуватися зі стінами або потрапляти у локальні мінімуми, оскільки карта зайнятості не враховується нейронною мережею. Крім того, SARL діє як локальний планувальник, оптимізуючи рух лише на невелику відстань вперед без інтеграції з глобальним маршрутом, що знижує ефективність навігації у великих просторах.

Суттєвим обмеженням методу є його детермінованість та нездатність враховувати мультимодальність людської поведінки. Генеруючи єдину найбільш імовірну стратегію руху, алгоритм не моделює ймовірнісний розподіл майбутніх траєкторій, що робить його вразливим у ситуаціях, коли людина різко змінює напрямок руху. Також впровадження методу ускладнюється значною ресурсоемністю механізму уваги. Обчислювальна складність алгоритму може зростати квадратично залежно від кількості людей. У соціальному середовищі це може призвести до критичних затримок у процесі прийняття рішень.

Порівняльна характеристика методів навігації на основі DRL у динамічному середовищі представлено в таблиці 1. 2.

Окрему групу сучасних підходів становлять методи прогнозування траєкторій пішоходів, які не здійснюють навігацію безпосередньо, але є ключовими допоміжними модулями для соціально-адаптивних навігаційних систем.

Фундаментальне значення для еволюції методів навігації має робота А. Алахі, в якій запропоновано архітектуру Social-LSTM [63]. Даний метод дозволив ефективно застосувати глибинне навчання для прогнозування поведінки. До його

появи домінували моделі, засновані на ручних функціях, такі як модель соціальних сил (Social Force Model, SFM), які часто не могли охопити складні, неявні правила людської взаємодії, особливо в щільних середовищах [64].

Таблиця 1. 2.

Порівняльна характеристика методів навігації на основі DRL у динамічному середовищі

Метод	Базова архітектура	Механізм соціальної взаємодії	Переваги	Недоліки
CADRL	DRL (Value Network)	Припущення взаємного уникнення зіткнень (Reciprocal Collision Avoidance)	Ефективне уникнення зіткнень у простих сценаріях.	Обмежена масштабованість при великій кількості агентів; обмежене врахування соціального контексту
LSTM-RL (GA3C-CADRL)	DRL (GA3C / A3C) + LSTM	Послідовне кодування станів динамічних агентів	Обробка змінної кількості агентів; врахування часових залежностей.	Виникнення проблеми «застиглого робота» у складних ситуаціях; обмежене моделювання соціальної взаємодії
SARL	DRL + LSTM + Attention	Social Attention (зважування важливості агентів)	Визначення найбільш релевантних агентів; висока ефективність у натовпі.	Неявне моделювання соціальної поведінки; обмежена інтерпретованість політики; зростання обчислювальної складності при великій кількості агентів

Ключовою ідеєю Social-LSTM є моделювання траєкторії кожної людини окремою нейронною мережею LSTM. Однак, оскільки рух людей є взаємозалежним (людина змінює траєкторію, реагуючи на сусідів), незалежні LSTM не здатні передбачити зіткнення. Для вирішення цієї проблеми було впроваджено шар «соціального пулінгу» (Social Pooling layer).

Механізм роботи Social-LSTM складається з таких етапів:

- кожен пішохід i у сцені використовує окремий екземпляр мережі LSTM для кодування історії власних координат (x, y) ;
- шар соціального пулінгу створює сітку навколо кожного агента для обміну інформацією між сусідніми об'єктами;

- приховані стани мереж LSTM сусідніх агентів розміщуються у відповідних клітинках сформованої сітки;
- оновлення стану кожного агента відбувається з урахуванням його власної історії та пулінгового вектора зі стиснутою інформацією про динаміку сусідів.

Метод Social-LSTM дозволив роботам «розуміти» не лише фізику руху, а й соціальні правила (наприклад, формування груп, обхід справа), навчаючись безпосередньо з даних реальних пішоходів, а не через запрограмовані формули сил.

Недоліком Social-LSTM є те, що він використовує фіксовану сітку для пулінгу, що є обчислювально затратним і локально обмеженим. Крім того, він генерує детермінований прогноз (одну найбільш імовірну траєкторію), ігноруючи мультимодальну природу людських рішень.

Хоча Social-LSTM спочатку був розроблений як метод прогнозування траєкторій, його ідеї еволюціонували у використанні генеративних моделей, таких як Social-GAN [65] та Social-BiGAT [66], у поєднанні з навчанням з підкріпленням. Ці методи вирішують проблему мультимодальності людської поведінки – факту, що з однієї точки людина може піти різними шляхами.

Метод Social-GAN (Social Generative Adversarial Networks) являє собою фундаментальний підхід до прогнозування руху пішоходів, який вперше успішно застосував архітектуру генеративно-змагальних мереж (GAN) для вирішення проблеми мультимодальності людської поведінки. Основна гіпотеза авторів полягає в тому, що методи на основі рекурентних мереж (такі як Social-LSTM), які оптимізуються за допомогою мінімізації середньоквадратичної похибки, схильні до усереднення прогнозів. Це призводить до генерації траєкторій, які є «середніми» математично, але фізично нереалістичними або соціально неприйнятними. Наприклад, рух крізь перешкоду замість її обходу. Social-GAN пропонує альтернативний підхід, де модель вчиться генерувати набір різноманітних, але правдоподібних варіантів майбутнього руху.

Архітектура методу складається з трьох ключових компонентів: генератора, дискримінатора та модуля соціального пулінгу. Генератор, побудований на базі

LSTM (Encoder-Decoder), приймає на вхід історію руху людей та вектор випадкового шуму, формуючи гіпотетичні майбутні траєкторії. Дискримінатор, також реалізований як LSTM, аналізує згенеровані шляхи разом із реальними даними та намагається відрізнити несправжні траєкторії від справжніх. У процесі змагального навчання генератор вчиться створювати реалістичні прогнози. У результаті синтезовані траєкторії набувають властивостей соціальної релевантності.

Вагомою інновацією Social-GAN є впровадження нового механізму агрегації контексту – Pooling Module. Він замінив сітку, яка використовувалася в Social-LSTM. Замість локального розбиття простору, Social-GAN застосовує глобальний пулінг. Відносні координати та приховані стани всіх агентів обробляються багатошаровим перцептроном, після чого застосовується симетрична функція (Max-Pooling) для отримання єдиного вектора взаємодії. Це дозволяє враховувати вплив усіх учасників сцени незалежно від їх розташування, значно знижуючи при цьому обчислювальну складність алгоритму.

З метою врахування мультимодальної природи прогнозів та забезпечення їхньої варіативності було впроваджено спеціалізовану функцію втрат – Variety Loss. На відміну від класичного підходу, в якому модель штрафується за відхилення всіх згенерованих варіантів від дійсних, Variety Loss заохочує мережу розподіляти прогнози, покриваючи різні ймовірні моди, напрямки руху. Під час навчання штрафується лише той варіант згенерованої траєкторії, який є найближчим до реальної, дозволяючи іншим варіантам досліджувати альтернативні шляхи без штрафів з боку функції втрат.

Попри значні переваги у моделюванні соціальних норм та мультимодальності, метод має певні недоліки. Як і більшість методів на базі GAN, Social-GAN є складним у навчанні через нестабільність збіжності генератора та дискримінатора. Крім того, оскільки метод фокусується на прогнозуванні, а не на керуванні, інтеграція його в систему навігації робота вимагає додаткових обчислювальних ресурсів для постійного перерахунку прогнозів та їх перетворення

в команди керування. Також базова версія методу не враховує статичні перешкоди сцени і карту приміщення, покладаючись лише на траєкторії динамічних об'єктів.

Наступним кроком у вдосконаленні генеративних підходів, спрямованим на поглиблене врахування структури соціальних взаємодій та підвищення стабільності моделювання, став метод Social-BiGAT.

Метод Social-BiGAT є генеративним підходом до прогнозування траєкторій руху пішоходів, який поєднує в собі механізми графічних мереж уваги (Graph Attention Networks, GAT) та архітектуру генеративно-змагальних мереж типу Bicycle-GAN. Головною метою методу є вирішення проблеми мультимодальності людської поведінки, тобто здатності моделювати різні ймовірні варіанти руху людини з однієї точки.

В основі архітектури Social-BiGAT лежить система, яка складається з генератора, двох типів дискримінаторів (локального та глобального) і латентного кодера.

Ключовим елементом моделювання соціальних взаємодій виступає графова мережа уваги (GAT). На заміну механізмів пулінгу або сортування за відстанню, Social-BiGAT представляє сцену як граф, де вузлами є люди, а ребрами – їхні взаємодії. Ваги цих ребер визначаються автоматично через механізм уваги, що дозволяє моделі динамічно оцінювати важливість кожного сусіднього агента незалежно від його фізичної відстані. Для врахування фізичного контексту середовища (наприклад, перешкод або дорожнього покриття) метод додатково використовує механізм м'якої уваги на основі зображення сцени, що обробляється згортковою нейронною мережею (CNN).

Перевагою Social-BiGAT є здатність генерувати реалістичні мультимодальні прогнози завдяки інтеграції принципів Bicycle-GAN. Метод створює двостороннє відображення між простором траєкторій та латентним простором шуму. Це досягається шляхом введення додаткового латентного кодера, який навчається відновлювати вектор шуму з генерованої траєкторії, що спонукає модель уникати проблеми «колапсу мод» і формувати різноманітні, але соціально прийнятні варіанти майбутнього руху.

Водночас, архітектурна складність методу може розглядатися як недолік у контексті обчислювальних витрат. Процес навчання Social-BiGAT є багатоступеневим і вимагає одночасної оптимізації чотирьох нейронних мереж із використанням комбінації п'яти різних функцій втрат, включаючи змагальні втрати (GAN loss), втрати реконструкції траєкторії (L2) та дивергенцію Кульбака-Лейблера для латентного простору. Така структура вимагає значних обчислювальних ресурсів для тренування та налаштування гіперпараметрів порівняно з простішими детермінованими моделями. Крім того, метод покладається на наявність візуальних даних для врахування фізичного контексту. Але це може бути обмеженням для роботів, оснащених лише лідарами.

Таблиця 1. 3.

Порівняльний аналіз методів прогнозування траєкторій

Метод	Базова архітектура	Механізм соціальної взаємодії	Тип прогнозу	Переваги	Недоліки
Social-LSTM	LSTM	Social Pooling (сітка навколо агента)	Детермінований	Враховує історію руху кожного пішохода; моделює локальні взаємодії.	Обчислювально затратний через фіксовану сітку; ігнорує глобальний контекст.
Social-GAN	GAN (Generator-Discriminator)	Global Pooling (узагальнення)	Стохастичний Мультимодальний	Генерує набір різноманітних, соціально прийнятних шляхів.	Складність навчання; відсутність явної оцінки ймовірності (потребує семплювання).
Social-BiGAT	GAN + GAT (Graph Attention)	Graph Neural Networks (соціальний граф)	Стохастичний	Найкраще моделює глобальні взаємозв'язки між усіма учасниками руху.	Висока обчислювальна складність; складність застосування в реальному часі.

Таким чином, проведений аналіз сучасних підходів до соціально-адаптивної навігації свідчить про поступову еволюцію від детермінованих фізичних моделей до гнучких інтелектуальних архітектур на основі глибинного навчання.

Незважаючи на переваги методів DRL, критичними залишаються питання стабільності процесу навчання у багатовимірних просторах станів та проблема «застиглого робота» у динамічних середовищах. Більшість існуючих рішень фокусується або на точності прогнозування, або на швидкодії керування, часто ігноруючи необхідність адаптивного балансу між безпекою та ефективністю досягнення цілі в умовах невизначеності.

Порівняльний аналіз методів прогнозування траєкторій представлено в таблиці 1. 3.

Виявлені обмеження обґрунтовують доцільність формування гібридних підходів, що синтезують можливості глибинного навчання з підкріпленням із сучасними стратегіями прискорення збіжності та адаптивними механізмами формування винагороди.

1.3. Постановка завдання та мети дослідження

Проведений у розділі 1 аналіз особливостей функціонування мобільних роботів у динамічних середовищах дозволив зробити висновки, що інтеграція роботизованих систем у спільний з людьми простір вимагає перегляду традиційних підходів до навігації. Класичні алгоритми планування траєкторії характеризуються високою ефективністю у статичних середовищах, проте в умовах динамічної взаємодії з рухомими перешкодами їх продуктивність суттєво знижується.

Водночас, огляд сучасних методів соціально-адаптивної навігації засвідчив перспективність використання підходів глибинного навчання з підкріпленням (DRL) та рекурентних нейронних мереж. Проте існуючі рішення демонструють обмежену здатність до інтерпретації стохастичності траєкторій людей.

Метою дослідження є підвищення ефективності та безпеки навігації автономних мобільних роботів у динамічних соціальних середовищах шляхом розробки методу адаптивного керування на основі глибинного навчання з підкріпленням.

Для досягнення поставленої мети необхідно вирішити такі наукові завдання:

1. Проаналізувати існуючі підходи до навігації мобільних роботів у соціальному середовищі та виявити їх переваги та недоліки.
2. Розробити математичну модель соціально-адаптивної навігації на основі марковського процесу прийняття рішень та формалізувати взаємодію робота з динамічними об'єктами.
3. Розробити метод навчання автономного мобільного робота на основі стратегії Curriculum Learning, який передбачає декомпозицію навчального процесу через поетапне ускладнення тренувальних сценаріїв у поєднанні з модифікацією структури функції винагороди.
4. Розробити архітектуру інтелектуального агента на основі алгоритму DRL для кількісної оцінки невизначеності середовища.
5. Удосконалити механізм адаптивного формування функції винагороди шляхом впровадження динамічного зважування компонентів функції винагороди.
6. Здійснити експериментальне дослідження ефективності розроблених методів у симуляційному середовищі.

Виконання визначених завдань забезпечить формування цілісного науково-методичного апарату для побудови інтелектуальних систем керування автономними роботами, здатними до проактивної соціальної адаптації в динамічному середовищі. Отримані результати дозволять розв'язати актуальну науково-практичну задачу підвищення автономності та безпеки мобільних платформ у динамічних середовищах.

Висновки до розділу 1

У першому розділі дисертаційної роботи проведено комплексний аналіз проблеми навігації автономних мобільних роботів у динамічних соціальних середовищах. Узагальнення результатів аналітичного огляду дозволяє сформулювати такі висновки:

1. Встановлено, що інтеграція автономних мобільних роботів у спільний з людьми простір вимагає переходу від критеріїв виключно фізичної безпеки до

забезпечення соціально прийнятної поведінки. Визначено, що ключовим викликом є стохастична природа руху людей, яка унеможлиблює використання детермінованих алгоритмів для забезпечення психологічного комфорту та передбачуваності маневрів робота.

2. З'ясовано, що класичні методи навігації ефективні лише в умовах повного спостереження. У сценаріях із рухом людей ці підходи демонструють критичні недоліки: нездатність моделювати складні патерни взаємодії людей, необхідність ручного налаштування параметрів та схильність до виникнення ефекту «застиглого робота».

3. Встановлено, що сучасні методи на основі глибинного навчання з підкріпленням забезпечують вищу адаптивність порівняно з класичними підходами. Проте більшість існуючих рішень використовують функції винагороди, які не враховують змінний рівень невизначеності середовища.

4. Послідовне вирішення окреслених завдань спрямоване на розробку методів соціально-адаптивної навігації. Запропонований концептуальний підхід дозволить подолати виявлені обмеження наявних систем, гарантуючи безпечне, ефективне та соціально комфортне функціонування АМР в умовах високої динамічної невизначеності.

РОЗДІЛ 2

ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ МОДЕЛЮВАННЯ РУХУ АВТОНОМНОГО МОБІЛЬНОГО РОБОТА

2.1. Концептуальні засади соціально-адаптивної навігації та формалізація проксемічних обмежень

За умов функціонування АМР в динамічному соціальному середовищі пріоритетним завданням навігації є гарантування безпечного, ефективного та соціально прийнятного переміщення АМР у спільному з людьми просторі. Складність вирішення цієї проблеми зумовлена необхідністю формалізації неявних соціальних норм та поведінкових патернів, яких люди дотримуються інтуїтивно. Ці аспекти важко піддаються опису у вигляді детермінованих правил або чітких метрик [67].

Необхідність математичного та алгоритмічного моделювання поведінки мобільного робота у соціальному середовищі зумовило виникнення таких термінологічних визначень, як «соціальна навігація» (social navigation), «соціально усвідомлена навігація» (socially aware navigation) та «соціально-адаптивна навігація» (socially adaptive navigation).

Семантичний аналіз цих понять дозволяє виявити суттєві відмінності в підходах до проектування систем керування. Термін «соціальна навігація» акцентує увагу на інтеграції робота в соціум з використанням правил, характерних для міжлюдської взаємодії. Втім, використання цього терміну несе ризики методологічної некоректності, створюючи хибне уявлення про тотожність когнітивних та соціальних можливостей АМР і людини.

У сучасних дослідженнях поняття «соціально усвідомлена навігація» [68] та «соціально-адаптивна навігація» [69] часто розглядаються як синонімічні. Спільним для обох визначень є вимога до здатності робота враховувати під час планування маневрів соціальний контекст, неявні правила людської поведінки та психологічний комфорт людей, не вимагаючи при цьому повної імітації людської

поведінки. Зазначені терміни охоплюють ідентичний набір критеріїв якості функціонування системи: гарантування безпеки, забезпечення соціальної прийнятності дій робота та передбачуваності його траєкторії руху.

У межах даного дисертаційного дослідження за базовий прийнято термін «соціально-адаптивна навігація». Такий вибір обґрунтовується тим, що поняття найбільш точно відображає фокус запропонованого підходу – здатність системи до динамічної адаптації навігаційної політики у відповідь на зміни сценаріїв взаємодії. Обраний термін підкреслює, що система не просто пасивно «усвідомлює» правила, а й активно пристосовує параметри руху до поточної ситуації.

Формалізація вимог дозволяє визначити критерії, за яких АМР класифікується як соціально-адаптивний [70]:

- АМР здатен детектувати людей, класифікувати їх як пріоритетні динамічні перешкоди та розглядати їхню фізичну безпеку як домінуючий фактор у системі прийняття рішень;
- у випадках навігаційних конфліктів (наприклад, перетин траєкторій) агент аналізує контекст і обирає стратегію (зупинка, об’їзд, зміна швидкості), яка відповідає очікуванням людей;
- поведінка АМР оптимізована для створення психологічного комфорту людини.

Враховуючи наведені критерії, ключовим параметром безпечної навігації виступає відстань фізичної взаємодії між АМР та людиною, зумовлюючи доцільність практичного впровадження моделі проксеміки в робототехнічні системи.

Проксеміка, як галузь соціальної психології, досліджує закономірності використання людиною фізичного простору в процесі комунікації. Фундаментальну концепцію зонування простору запропонував Едвард Т. Холл, виділивши чотири простори: інтимний, персональний, соціальний та публічний [71].

Визначимо режими функціонування АМР в кожному проксемічному просторі.

Перебування робота в інтимному просторі ($r < 0,5$ м) викликає максимальний психологічний дискомфорт і розцінюється як загроза фізичній безпеці людини. Дія робота – екстренне гальмування. Навігаційна стратегія – повна зупинка руху до моменту збільшення дистанції або зміни траєкторії руху людиною.

Особистий (персональний) простір ($0,5 \text{ м} < r < 1,2 \text{ м}$) призначений для комфортної взаємодії людей. Дія робота – максимальне зниження швидкості. Навігаційна стратегія – рух робота в межах особистого простору дозволяється виключно за умови низької швидкості та наявності чіткої мети комунікації. За ініціативи людини робот може перебувати в цьому просторі для виконання сервісних функцій (передача предмета, зчитування інформації).

Соціальний простір ($1,2 \text{ м} < r < 3,5 \text{ м}$) є простором для формального спілкування і вважається оптимальною дистанцією для виконання маневрів роботом. Дія робота – рух зі зниженою швидкістю. Навігаційна стратегія – робот здійснює обхід людини, мінімізуючи різкі зміни курсу. Швидкість руху обмежується для забезпечення передбачуваності дій робота.

Публічний простір ($r > 3,5 \text{ м}$) характерний для взаємодії незнайомих людей або публічних виступів. Вплив присутності робота на психоемоційний стан людини в цьому просторі є мінімальним. Дія робота – вільне переміщення. Навігаційна стратегія – планування траєкторії здійснюється без соціальних обмежень, з пріоритетом на швидкість досягнення цілі.

Систематизоване представлення проксемічних зон та відповідних стратегій обмеження швидкості руху агента наведено на рис. 2.1.

Варто зазначити, що розміри проксемічних зон не є статичними константами і демонструють залежність від ситуативного контексту та щільності натовпу. В умовах обмеженого середовища (ліфти, громадський транспорт, вузькі коридори) спостерігається явище вимушеного стиснення зон, коли порушення особистого простору сприймається людиною комфортно [72].

Таким чином, радіус публічного простору може ситуативно зменшуватися до розмірів інтимного простору. Варіативність залежить як від індивідуальних психологічних особливостей людини, так і від структурних особливостей

культурного середовища, охоплюючи етнічні традиції просторової взаємодії, прийняті норми невербальної комунікації та поділ суспільств на висококонтактні й низькоконтактні.

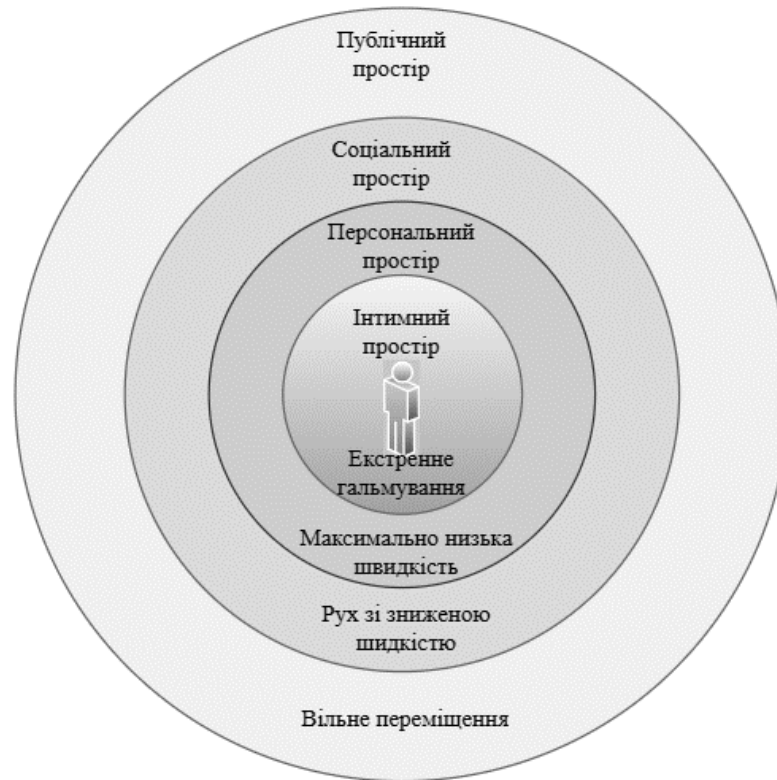


Рис. 2.1 Модель проксемічних зон та відповідні обмеження навігаційної поведінки робота

Окрім радіальної відстані, суттєве значення має анізотропія особистого простору. Дослідження підтверджують, що люди більш чутливі до наближення у фронтальній зоні (попереду), вважаючи рух зустрічного об'єкта найбільш дискомфортним. Тоді як наближення збоку або позаду сприймається менш критично [73].

Узагальнюючи викладене, можна стверджувати, що проксемічні правила та зони комфорту визначають систему просторових обмежень, необхідних для формування соціально-адаптивної навігаційної поведінки мобільного робота.

2.2. Моделювання навігації АМР у динамічному середовищі

Математичною основою для розв'язання задач послідовного прийняття рішень, де АМР для досягнення мети взаємодіє із динамічним середовищем, є апарат марковських процесів прийняття рішень (Markov Decision Process, MDP).

Формально задачу навігації можна описати кортежем з п'яти елементів (S, A, P, R, γ) , де S – простір станів, що описує множину всіх можливих конфігурацій середовища та агента. Стан $s_t \in S$ у момент часу t містить інформацію, необхідну для прийняття рішення (наприклад, дані лазерного далекоміра, координати цілі, кінематичні параметри робота); A – простір дій, який визначає множину доступних команд керування. Дія $a_t \in A$ – це вектор керування, застосований агентом у момент часу t ; $P(s'|s, a)$ – функція ймовірності переходу, що визначає розподіл ймовірностей переходу в новий стан s' за умови виконання дії a у стані s . Функція моделює динаміку середовища, враховуючи як кінематику робота, так і стохастичні фактори (ковзання коліс, непередбачуваний рух динамічних об'єктів); $R(s, a)$ – функція винагороди, яка зіставляє скалярне значення (підкріплення) кожній парі «стан-дія». Дана функція є важливим елементом системи, оскільки кількісно визначає мету навчання: позитивні значення стимулюють наближення до цілі, а негативні (штрафи) – запобігають зіткненням та небезпечним маневрам; $\gamma \in [0, 1)$ – коефіцієнт дисконтування, який регулює пріоритетність майбутніх винагород відносно миттєвих. Значення $\gamma \rightarrow 1$ орієнтує агента на довгострокове планування траєкторії.

У контексті автономної навігації простір станів S формується на основі локальних спостережень. Стан s_t може бути представлений як вектор, який включає лінійну v та кутову w швидкості, відносні координати цілі та масив даних сенсорів.

Простір дій A для робота з диференціальним приводом, як правило, є неперервним і складається з пари $a_t = [\vartheta_t, \omega_t]$. При формалізації простору дій A необхідно враховувати кінематичні обмеження мобільної платформи

$$A = \{(v, \omega) \in R^2 | 0 \leq v \leq v_{max}, |\omega| \leq \omega_{max}\}. \quad (2.1)$$

Врахування цих обмежень на рівні MDP дозволяє генерувати фізично можливі траєкторії для виконавчих механізмів робота.

Визначені простори станів S та дій A , разом із функцією винагороди R , дозволяють кількісно описати мету керування – безпечний рух до цілі з уникненням перешкод. Такий підхід дає змогу звести задачу навігації до оптимізаційної задачі максимізації очікуваної винагороди, що обґрунтовує доцільність використання алгоритмів глибинного навчання з підкріпленням (DRL) для синтезу адаптивних стратегій керування мобільним роботом.

2.3. Теоретичні основи методів глибинного навчання з підкріпленням

Сучасні технології Deep Learning (DL) і Reinforcement Learning (RL) стали ефективними методами для вирішення задач планування руху мобільного робота у невідомому динамічному середовищі, зокрема й соціальному [74].

Навчання з підкріпленням використовується для вирішення задач послідовного прийняття рішень. Базуючись на принципі проб і помилок, технологія RL використовує функцію винагороди, отриману від взаємодії робота з навколишнім середовищем, як сигнал зворотного зв'язку для навчання АМР. Сигнали підкріплення надаються середовищем і використовуються для оцінки виконаних дій [75].

Навчання з підкріпленням використовується у робототехніці для створення карт середовищ, для планування шляху роботів (уникнення перешкод, здійснення маневрів, вибір швидкості та напрямку руху), для визначення оптимального використання ресурсів [76].

Ключовими компонентами RL є:

- 1) агент, який приймає рішення;
- 2) середовище, з яким постійно взаємодіє агент;
- 3) стан s_t відображає перебування агента в середовищі в момент часу t ;
- 4) дія агента a_t , обрана агентом у стані s_t ;

- 5) винагорода r_t , надана середовищем у відповідь на виконану дію;
- 6) політика π , яка регламентує поведінку агента в середовищі.

Політика агента π – це відображення простору станів у простір дій:

$$\pi: S \rightarrow A$$

У стані $s_t \in S$, агент виконує дію $a_t \in A$, переходить до наступного стану s_{t+1} відповідно до ймовірності переходу стану P , одночасно отримуючи винагороду $r_t \in R$ від середовища. Політика π обчислює ймовірності вибору дій $a \in A$:

$$\sum_{a \in A} \pi(s, a) = 1. \quad (2.2)$$

У процесі навчання робот постійно взаємодіє з середовищем для формування оптимальної політики, орієнтованої на максимізацію довгострокової винагороди [77].

Використавши коефіцієнт $\gamma \in [0, 1)$ можна обчислити значення кумулятивної винагороди таким чином:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}. \quad (2.3)$$

Функція цінності стану s_t визначається через $V_\pi(s)$, а функція цінності пари стан-дія – $Q_\pi(s, a)$. Ці значення використовуються для оцінки довгострокової винагороди, очікуваної агентом за умови слідування політиці π .

$$V_\pi(s) = E_\pi[R_t | s_t = s], \quad (2.4)$$

$$Q_\pi(s, a) = E_\pi[R_t | s_t = s, a_t = a]. \quad (2.5)$$

$V_\pi(s)$ і $Q_\pi(s, a)$ можна виразити в рекурсивній формі для встановлення зв'язку між станами $s = s_t$ і $s' = s_{t+1}$

$$V_\pi(s) = \sum_a \pi(s, a) \sum_{s'} P_{ss'}^a [R_{ss'}^a + \gamma V_\pi(s')], \quad (2.6)$$

$$Q_\pi(s, a) = \sum_{s'} P_{ss'}^a [R_{ss'}^a + \gamma \sum_{a'} Q_\pi(s', a')], \quad (2.7)$$

де $P_{ss'}^a = P(s_{t+1}|s_t = s, a_t = a)$, $R_{ss'}^a = E(r_{t+1}|s_t, s_{t+1}, a_t)$.

Рівняння (2.6) та (2.7) відомі як рівняння Беллмана, отримані за допомогою динамічного програмування шляхом оптимізації функції цінності [78].

Роботизовані системи на основі RL неминує зустрічатися з дилемою дослідження та експлуатації. Агент повинен використовувати накопичені знання для отримання винагороди, одночасно досліджуючи невідоме середовище для вивчення нових стратегій та покращення вибору дій у майбутньому [79].

Ефективне балансування між дослідженням нових стратегій та експлуатацією накопиченого досвіду потребує наявності математичного механізму оцінки якості прийнятих рішень. У контексті MDP такий механізм реалізується через ітеративне оновлення функцій цінності, дозволяючи кількісно визначати перспективність кожної дії в конкретному стані. Прикладом такого підходу вважається алгоритм Q-learning, який заклав фундамент для розвитку сучасних безмодельних методів керування.

На початковому етапі алгоритм Q-learning ініціалізує функцію $Q(s, a)$ та поточний стан s , після чого генерує дію a згідно з ϵ -жадібною політикою, зумовлюючи зміну стану системи на s' та отримання винагороди r [80]. Далі значення Q оновлюється відповідно до правила

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)]. \quad (2.8)$$

Двома основними параметрами в алгоритмі Q-learning є α і γ . Швидкість навчання α визначає ступінь впливу нової інформації на коригування попередніх значень. Низьке значення швидкості навчання свідчить про пріоритетність збереження накопиченого досвіду над інтеграцією нових даних. Коефіцієнт дисконтування γ визначає міру впливу майбутніх винагород на поточну оцінку політики [81].

Зазначений алгоритм використовує Q-таблицю, значення якої застосовуються для вибору оптимальної дії. Після досягнення агентом місця призначення, алгоритм завершує ітерацію, і агент повертається до початкового вузла для продовження навчального циклу [82].

Концептуально алгоритм Q-learning інтегрує принципи динамічного програмування та методів Монте-Карло для ітеративного розв'язання рівняння оптимальності Беллмана. Незважаючи на широке розповсюдження завдяки структурній простоті та високій конвергентній здатності, цей метод має низку обмежень у контексті складних навігаційних задач. Зокрема, поелементне оновлення значень для кожної пари «стан-дія» суттєво ускладнює апроксимацію оптимальної стратегії у просторах високої розмірності. Крім того, при масштабуванні середовища дискретна природа Q-таблиць зумовлює експоненціальне зростання обчислювальної пам'яті. Вказані недоліки обмежують практичне застосування алгоритму в задачах із неперервним простором станів [81].

Зазначені обмеження зумовили необхідність розробки нових підходів, які не потребують явного збереження повного простору станів середовища. Ефективним розв'язанням цієї проблеми стала інтеграція ітеративної логіки RL із можливостями DL. Запропонована парадигма, реалізована в методах DRL, забезпечила перехід від використання дискретних Q-таблиць до застосування нейромережових архітектур у ролі універсальних апроксиматорів [83].

Залежно від об'єкта апроксимації сучасні методи DRL поділяються на методи на основі функції цінності, методи на основі політики та гібридні підходи на основі архітектури актора-критика.

2.3.1. Методи на основі функції цінності

DRL з апроксимацією функції цінності опосередковано отримує політику агента шляхом повторного оновлення функції цінності. Після досягнення оптимального значення обчислюється оптимальна політика агента [84].

Практична реалізація таких методів у багатовимірних просторах станів потребує використання стабільних механізмів апроксимації, здатних нівелювати затримки в отриманні винагороди. Архітектурою, яка успішно вирішила ці виклики шляхом інтеграції глибоких нейронних мереж у процес Q-навчання, став алгоритм DQN (Deep Q-Network, DQN). У DQN використано згорткову нейронну мережу для представлення функції цінності дії.

Даний алгоритм є різновидом класичного алгоритму Q-learning з такими основними доповненнями:

- застосовано архітектуру глибинної згорткової нейронної мережі для апроксимації Q-функції;
- використано мініпакекти випадкових тренувальних даних замість однокрокових оновлень досвіду;
- використано попередні параметри мережі для оцінки Q-значень наступного стану.

У алгоритмі DQN зберігається велика кількість останніх подій, де кожна подія містить кортеж з п'яти складових (s, a, r, s', T) . Агент виконує дію a у стані s , переходить в стан s' і отримує винагороду r ; T є логічним значенням, що вказує, чи s' є кінцевим станом. Після кожного кроку в середовищі агент додає досвід до своєї пам'яті. Після деякої кількості кроків агент випадковим чином вибирає міні-пакет зі своєї пам'яті для виконання оновлень Q-функції. Повторне використання попереднього досвіду для оновлення Q-функції відоме як відтворення досвіду [85].

Основна перевага архітектури DQN полягає у використанні глибинних нейронних мереж як універсальних апроксиматорів, що дозволяє ефективно знижувати розмірність вхідних даних та формувати наближені значення Q-функції [86]. Водночас алгоритм має недоліки: низька швидкість збіжності, висока дисперсія результатів та схильність до систематичного завищення оцінок цінності дій. Крім того, орієнтованість DQN виключно на дискретний простір керування зумовлює переривчастий характер руху робота і знижує якість навігації в реальних умовах [87].

Для подолання проблеми переоцінки цільових значень та підвищення стабільності навчання Ван Хасселтом та ін. було розроблено вдосконалену модифікацію методу – алгоритм Double DQN (DDQN) [88]. Основна ідея цього підходу полягає у розділенні процесів вибору дії та її оцінювання.

Архітектура DDQN базується на використанні двох нейронних мереж із незалежними просторами вагових коефіцієнтів. Основна мережа здійснює вибір оптимальної дії з множини доступних варіантів, після чого цільова мережа виконує

оцінювання обраної дії для обчислення її Q-значення. Процес навчання передбачає ітеративне оновлення параметрів основної мережі на кожному кроці, тоді як вагові коефіцієнти цільової мережі періодично синхронізуються шляхом копіювання параметрів основної моделі. Рівняння Беллмана в цьому алгоритмі має вигляд [89]:

$$Q(s, a, \theta) = r + \gamma Q(s', \operatorname{argmax}_{a'} Q(s', a'; \theta); \theta'). \quad (2.9)$$

Незважаючи на наявність двох мереж, обчислювальна складність DDQN залишається низькою, що дозволяє використовувати його на вбудованих платформах мобільних роботів. Також DDQN демонструє кращу здатність до узагальнення порівняно з базовим DQN, особливо за наявності шумів сенсорів у вхідних даних.

Застосування DDQN у задачах соціальної навігації обмежена низькою чинників: дискретність керування; ігнорування соціального контексту при ідентифікації динамічних агентів; ризик виникнення ефекту «застиглого робота».

2.3.2. Методи на основі політики

Методи DRL з оптимізацією політики дозволяють ефективно навчати AMP у складних середовищах, в яких необхідно приймати рішення в реальному часі [83]. Вони обчислюють градієнт очікуваної винагороди відносно параметрів політики і оновлюють їх у напрямку збільшення очікуваної винагороди.

Замість використання функції цінності для оцінки станів або дій, ці методи безпосередньо оптимізують політику π . Такий підхід забезпечує можливість роботи з неперервними просторами дій, важливими для плавного керування швидкістю мобільного робота в соціальному середовищі.

Представником цієї групи є метод Reinforce, математична реалізація якого ґрунтується на стохастичній апроксимації за схемою Монте-Карло [84]. Процес навчання передбачає збір повного епізоду взаємодії AMP із середовищем перед виконанням кроку оновлення вагових коефіцієнтів моделі. Для оцінювання градієнта застосовується фактична кумулятивна винагорода завершеного епізоду.

Такий підхід дозволяє системі керування ефективно навчатися без побудови апріорної математичної моделі динаміки соціального середовища.

Застосування алгоритмів на основі оптимізації політики для задач навігації мобільних роботів характеризуються такими перевагами:

- здатність безпосередньо працювати з багатовимірними та неперервними просторами дій без необхідності аналітичного пошуку максимального значення функції цінності на кожному кроці;
- можливість формування стохастичних політик для забезпечення балансу між активним дослідженням невідомих просторів середовища та використанням попередньо набутого навігаційного досвіду;
- забезпечення стабільної збіжності процесу навчання завдяки прямій та безперервній оптимізації параметрів цільової функції.

Водночас базові алгоритми градієнта політики мають суттєві обмеження. Застосування методів Монте-Карло супроводжується значною дисперсією оцінок градієнта внаслідок через значну варіативність траєкторій та послідовностей станів у різних епізодах. Алгоритми характеризуються низькою ефективністю використання вибірки, потребуючи генерації нових наборів даних після кожного кроку оновлення параметрів нейронної мережі. Крім того, потреба очікування завершення епізоду моделювання уповільнює процес пошуку оптимального рішення у задачах соціальної навігації з тривалим часовим горизонтом.

2.3.3. Методи актора–критика

Для подолання вищезазначених обмежень виникла об'єктивна потреба у заміні емпіричних оцінок Монте-Карло на стабільніші наближення на основі методів часової різниці (Temporal Difference). Забезпечення можливості покрокового оновлення досягається шляхом оцінювання поточного стану агента безпосередньо під час виконання завдання, не чекаючи термінального стану. Такий підхід вимагає залучення додаткового обчислювального компонента, здатного апроксимувати очікувану винагороду та коригувати процес пошуку оптимальної траєкторії.

Практичною реалізацією цього підходу є гібридна архітектура актора-критика (Actor-Critic, AC), яка інтегрує переваги методів на основі політики та функцій цінності [90]. Актор є функцією політики, яка відповідає за створення дій і взаємодію з середовищем. Критик використовує функцію цінності для оцінки продуктивності та керування діями актора на наступному етапі [91].

Незважаючи на високу ефективність архітектури актора-критика в умовах неперервного простору дій, її послідовний характер зумовлює виражену кореляційну залежність між суміжними вибірками досвіду, дестабілізуючи процес градієнтного спуску. Для подолання залежності без використання енерговитратних буферів відтворення було розроблено концепцію паралельного навчання. Такий підхід дозволив трансформувати стандартну модель у асинхронний метод актора-критика із функцією переваги (Asynchronous Advantage Actor-Critic, A3C) [92].

A3C використовує одночасну взаємодію декількох екземплярів актора з різними копіями середовища, усуваючи кореляційну залежність даних та суттєво прискорюючи збіжність алгоритму [93].

Базуючись на структурі AC, A3C вносить такі вдосконалення:

1. Асинхронне розпаралелювання навчання шляхом ініціалізації множини незалежних середовищ, де паралельні агенти з локальними копіями мереж досліджують простір станів та незалежно оновлюють параметри глобальної моделі.
2. Оновлення функції цінності критика на основі багатокрокової кумулятивної винагороди для більш ефективного ітераційного поширення оцінок та стрімкого зростання швидкості збіжності алгоритму.

Незважаючи на складність налаштування гіперпараметрів, алгоритм A3C демонструє вищу обчислювальну ефективність порівняно з архітектурою DQN, що робить його придатним для завдань навігації в реальному часі. Проте асинхронна природа оновлень у A3C іноді призводить до нестабільності навчання через значні відхилення параметрів локальних агентів від глобальної мережі.

Для подолання цього недоліку було розроблено алгоритм A2C (Advantage Actor-Critic, A2C), який є синхронною модифікацією попереднього методу [94].

Головна відмінність A2C полягає в тому, що він очікує завершення кроків усіма паралельними агентами перед виконанням чергового градієнтного оновлення. Такий синхронний підхід забезпечує більшу стабільність процесу навчання та ефективне використання обчислювальних ресурсів. Перевагами A2C є висока ефективність використання вибірки, стабільність збіжності та адаптованість до дискретних і неперервних просторів дій, що дозволяє успішно застосовувати його в задачах керування AMP [95].

Незважаючи на стабільність синхронних оновлень у A2C, традиційні методи актора-критика часто схильні до швидкої втрати різноманітності поведінки, що обмежує здатність агента до ефективного дослідження складних середовищ. З метою усунення цієї проблеми було розроблено алгоритм SAC (Soft Actor-Critic, SAC), який базується на концепції максимізації ентропії [96].

На відміну від попередніх підходів, алгоритм SAC оптимізує не лише очікувану кумулятивну винагороду, а й ступінь стохастичності дій агента. Застосування цього механізму забезпечує високу робастність навчання, успішно запобігаючи передчасній збіжності до локальних оптимумів у задачах із неперервним простором керування. SAC інтегрує максимізацію ентропії в своїй архітектурі і досягає балансу між дослідженням та експлуатацією, роблячи навчання більш стабільним та ефективним [97].

Практична реалізація алгоритму SAC передбачає використання ансамблю з п'яти нейронних мереж:

- мережа політики $\pi_{\phi}(a_t|s_t)$ для генерації дій агента;
- мережа функції цінності стану $V_{\psi}(s_t)$ для оцінювання поточного стану;
- цільова мережа функції цінності стану $V_{\bar{\psi}}(s_t)$ для стабілізації процесу навчання;
- дві незалежні мережі м'якої Q-функції $Q_{\theta_1}(s_t, a_t)$ і $Q_{\theta_2}(s_t, a_t)$ для запобігання переоцінці цільових значень.

Таким чином, цільова функція функції цінності м'якого стану визначається як

$$J_V(\psi) = E_{s_t \sim D} \left[\frac{1}{2} (V_\psi(s_t) - E_{a_t \sim \pi_\varphi} [Q_\theta(s_t, a_t) - \log \pi_\varphi(a_t | s_t)])^2 \right]. \quad (2.10)$$

Гradient обчислюється як:

$$\hat{\nabla}_\psi J_V(\psi) = \nabla_\psi V_\psi(s_t) (V_\psi(s_t) - Q_\theta(s_t, a_t) + \log \pi_\varphi(a_t | s_t)). \quad (2.11)$$

Цільова функція функції м'якого значення Q визначається як:

$$J_Q(\theta) = E_{(s_t, a_t) \sim D} \left[\frac{1}{2} Q_\theta(s_t, a_t) - \hat{Q}(s_t, a_t)^2 \right], \quad (2.12)$$

де $\hat{Q}(s_t, a_t) = r(s_t, a_t) + \gamma E_{s_{t+1} \sim \rho} [V_{\bar{\psi}}(s_{t+1})]$

Gradient:

$$\hat{\nabla}_\theta J_Q(\theta) = \nabla_\theta Q_\theta(a_t, s_t) (Q_\theta(s_t, a_t) - r(s_t, a_t) - \gamma V_{\bar{\psi}}(s_{t+1})). \quad (2.13)$$

Цільова функція оновлення стратегії:

$$J_\pi(\varphi) = E_{s_t \sim D} \left[D_{KL}(\pi_\varphi(\cdot | s_t) || \frac{e^{Q_\theta(s_t, \cdot)}}{Z_\theta(s_t)}) \right]. \quad (2.14)$$

Дії вибираються за допомогою методу повторної параметризації

$$a_t = f_\varphi(\epsilon_t; s_t),$$

де ϵ_t – вектор вхідного шуму.

Алгоритм SAC має значну перевагу перед іншими алгоритмами навчання з підкріпленням завдяки його можливостям дослідження і здатності ефективно адаптуватися до більш складних завдань в середовищах з неперервним простором дій [98].

Альтернативою стохастичним підходам, орієнтованим на розширене дослідження середовища шляхом максимізації ентропії, є методи формування детермінованих стратегій. Розвиток цього напрямку дозволив адаптувати механізми стабілізації навчання, характерні для дискретних моделей, до задач із неперервним простором дій. Ці ідеї були втілені в алгоритмі глибинного детермінованого

градієнта політики (Deep Deterministic Policy Gradient, DDPG), створеного дослідниками Google DeepMind як функціональне розширення архітектури DQN [99].

З метою стабілізації обчислювального процесу архітектура DDPG інтегрує ключові механізми стабілізації, зокрема відтворення досвіду (Experience Replay) та цільові мережі (Target Networks).

Перший механізм реалізується шляхом формування мініпакетів випадкових вибірок із накопиченого буфера пам'яті, суттєво зменшуючи кореляцію між послідовними переходами (s, a, r, s') у середовищі.

Другим механізмом виступають цільові мережі, які зменшують нестабільність навчання, спричинену швидкими змінами цільових значень у процесі градієнтного спуску. На відміну від DQN-подібних підходів, DDPG застосовує стратегію м'якого оновлення (soft update), забезпечуючи ітеративне коригування параметрів цільових мереж.

У контексті навігації мобільних роботів алгоритм DDPG дозволяє формувати стратегію плавного руху без дискретних стрибків швидкостей [100].

Завдяки здатності ефективно оперувати у багатовимірних просторах станів та дій, DDPG демонструє високу результативність при вирішенні задач автономного керування у складних динамічних середовищах.

Попри високу результативність DDPG у неперервних середовищах, цей алгоритм залишається надмірно чутливим до вибору гіперпараметрів, а неконтрольовані різкі зміни в оновленні параметрів часто призводять до дестабілізації всієї моделі.

Прагнення поєднати гнучкість методів градієнта політики з надійністю довірчих областей (trust regions) зумовило появу алгоритму PPO (Proximal Policy Optimization, PPO).

Алгоритм PPO забезпечує стабільність навчання та високу ефективність використання даних, а також вирізняється здатністю до масштабування для роботи у складних динамічних середовищах [101].

Головною метою PPO є максимізація очікуваної винагороди $J(\pi_\theta)$ за умови, що оновлена політика не буде надмірно відхилятися від попередньої. Цільова функція оптимізації визначається формулою:

$$L_{CLIP}(\theta) = E_t[\min(r_t(\theta)A_t, \text{clip}(r_t(\theta), 1 - \varepsilon, 1 + \varepsilon)A_t)], \quad (2.15)$$

де $r_t(\theta)$ – відношення ймовірностей, визначається наступним чином:

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}, \quad (2.16)$$

де A_t – оцінка функції переваги;

ε – гіперпараметр, який задає межі діапазону відсікання;

$\text{clip}(\cdot)$ обмежує відношення ймовірностей інтервалом $[1 - \varepsilon, 1 + \varepsilon]$.

Функція переваги A_t для дії a_t у стані s_t розраховується за формулою

$$A_t = Q_\pi(s_t, a_t) - V_\pi(s_t), \quad (2.17)$$

де $Q_\pi(s_t, a_t)$ – функція цінності дії;

$V_\pi(s_t)$ – функція цінності стану.

Ключовою особливістю методу є забезпечення плавного оновлення параметрів за допомогою механізму обмеження (clipping) змін політики. Такий підхід дозволяє уникнути різких змін у стратегії агента, забезпечує стабільну збіжність алгоритму та підтримує ефективний баланс між дослідженням середовища та використанням набутого досвіду.

Крім компонента відсікання, повна функція втрат у PPO включає втрати функції цінності $L_V(\theta)$ та втрати ентропії $L_E(\theta)$:

$$L(\theta) = L_{CLIP}(\theta) - c_1 L_V(\theta) + c_2 L_E(\theta), \quad (2.18)$$

де c_1, c_2 – коефіцієнти регулювання вагового внеску втрат функції цінності та ентропії у загальну функцію втрат;

$L_V(\theta)$ – функція втрат, яка апроксимує функцію цінності $V(s)$:

$$L_V(\theta) = E_t[(V_\theta(s_t) - R_t)^2], \quad (2.19)$$

$L_E(\theta)$ – втрата ентропії, яка стимулює дослідження:

$$L_E(\theta) = E_t \left[- \sum_a \pi_\theta(a|s_t) \log \pi_\theta(a|s_t) \right]. \quad (2.20)$$

Порівняно з альтернативними методами глибинного навчання з підкріпленням, PPO демонструє значно меншу чутливість до вибору гіперпараметрів. Ця особливість суттєво спрощує налаштування моделі для специфічних навігаційних сценаріїв. Зазначена робастність забезпечується здатністю алгоритму підтримувати стабільну швидкість навчання без ризику дивергенції політики через невдалий вибір кроку градієнта.

Окрім цього, PPO функціонує як on-policy алгоритм із оновленням параметрів на основі даних поточної стратегії. Такий підхід гарантує пряму кореляцію між досвідом агента та траєкторією його оптимізації. Водночас це усуває помилки апроксимації, притаманні off-policy методам внаслідок використання застарілих даних із буфера відтворення.

2.3.4. Порівняльний аналіз алгоритмів DRL

Ефективна навігація автономних мобільних роботів у динамічних середовищах забезпечується завдяки стабільності навчання, ефективності використання даних, здатності працювати в безперервних просторах дій, а також адаптації до динамічного та частково невизначеного середовища із присутністю людей. Кожен із розглянутих алгоритмів характеризується перевагами та обмеженнями, які визначають доцільність їх використання в умовах соціальної навігації.

Обмеженням методів DQN та DDQN залишається їхня орієнтованість на дискретні простори керування. Оскільки задачі соціальної навігації вимагають безперервного контролю швидкості для збереження кінематичної плавності руху, застосування дискретизації зумовлює неприродну та соціально неприйнятну поведінку мобільного робота.

Алгоритми A2C та A3C підтримують як дискретні, так і безперервні простори дій. A3C використовує асинхронне навчання кількох агентів, яке сприяє швидшій

збіжності та кращому дослідженню середовища, зокрема в динамічних соціальних сценаріях. Водночас A2C є синхронною версією A3C, яка спрощує реалізацію та підвищує відтворюваність результатів. Недоліком обох алгоритмів є відносно висока варіативність навчання та чутливість до вибору гіперпараметрів. У контексті соціальної навігації це може проявлятися у нестабільній або непередбачуваній поведінці робота поблизу людей.

Незважаючи на здатність оперувати неперервними просторами керування, алгоритм DDPG виявляє низку фундаментальних недоліків у контексті соціально-адаптивної навігації. Здатність алгоритму до переоцінки функції цінності провокує агресивну поведінку робота в динамічному середовищі і підвищує ймовірність зіткнень. Висока обчислювальна нестабільність та гіперчутливість моделі до налаштувань спричиняють небезпечні осциляції керуючих сигналів.

Основною перевагою PPO є стабільність навчання, яка досягається за рахунок обмеження кроку оновлення політики. Алгоритм добре працює в безперервних просторах дій, є відносно нечутливим до гіперпараметрів і демонструє високу відтворюваність результатів.

У задачах соціальної навігації PPO добре поєднується з багатокomпонентними функціями винагороди, які враховують проксеміку, комфорт людей, уникнення зіткнень та ефективність руху.

SAC забезпечує баланс між дослідженням та експлуатацією, який критично важливий у складних соціальних середовищах. Алгоритм демонструє високу ефективність і здатність до узагальнення, а також стабільне навчання в безперервних просторах дій. Недоліком алгоритму є вища обчислювальна складність і складніша реалізація порівняно з PPO.

Узагальнене порівняння алгоритмів DRL наведено у таблиці 2. 1.

Результати проведеного порівняльного аналізу свідчать, що алгоритми, які функціонують у дискретних просторах дій, зокрема DQN та DDQN, мають обмежений потенціал застосування у задачах соціально-адаптивної навігації. Натомість методи архітектури актора-критика із підтримкою неперервного простору дій (DDPG, A2C, A3C, PPO, SAC) демонструють вищу релевантність

вимогам динамічних соціальних середовищ. Серед досліджених методів алгоритми PPO та SAC забезпечують оптимальний компроміс за критеріями стабільності збіжності, якості навігаційної поведінки, здатності до масштабування у складних сценаріях взаємодії.

Таблиця 2. 1.

Порівняльний аналіз методів DRL

Алгоритм	Простір дій	Тип політики	Стабільність навчання	Ефективність навчання	Придатність до соціальної навігації
DQN	Дискретний	Off-policy	Низька-середня	Середня	Низька
DDQN	Дискретний	Off-policy	Середня–висока	Середня	Низька
A2C	Дискретний/безперервний	On-policy	Середня-висока	Середня	Середня
A3C	Дискретний/безперервний	On-policy	Середня-висока	Середня–висока	Середня
DDPG	Безперервний	Off-policy	Низька	Середня–висока	Середня
PPO	Дискретний/безперервний	On-policy	Висока	Висока	Висока
SAC	Безперервний	Off-policy	Висока	Висока	Середня-висока

Враховуючи специфіку соціально-адаптованої навігації, як базовий метод для подальшого дослідження обрано алгоритм PPO. Такий вибір є обґрунтованим з огляду на специфіку задачі соціально-орієнтованої навігації AMP.

По-перше, завдяки механізму обмеження оновлення політики (clipping), PPO запобігає деструктивним змінам у вагових коефіцієнтах нейронної мережі. Висока стабільність навчання цього алгоритму дозволяє використовувати складні функції винагороди, які враховують соціальні норми, проксемічні обмеження та взаємодію з людьми. SAC, на відміну від PPO, орієнтований на максимізацію ентропії і заохочує стохастичність дій, що може призводити до надмірної варіативності траєкторій руху в соціальних середовищах.

По-друге, PPO демонструє високу відтворюваність результатів і меншу чутливість до налаштування гіперпараметрів порівняно з SAC.

По-третє, PPO є алгоритмом on-policy, що забезпечує прямий зв'язок між поточною політикою та зібраними траєкторіями. Це є принципово важливим у динамічному соціальному середовищі, де поведінка людей може змінюватися від епізоду до епізоду. На відміну від off-policy підходу SAC, який повторно використовує дані з буфера досвіду, PPO зменшує ризик навчання на неактуальних сценаріях взаємодії.

Таким чином, вибір алгоритму PPO у даному дослідженні зумовлений прагненням досягти балансу між стабільністю навчання, передбачуваністю соціальної поведінки і відтворюваністю експериментальних результатів. Подальші дослідження зосереджені на вдосконаленні алгоритму PPO шляхом адаптації архітектури нейронної мережі та механізмів навчання для підвищення ефективності соціально-орієнтованої навігації мобільного робота.

2.4. Модель соціально-адаптивної навігації мобільного робота

Методологія розробки та навчання архітектур соціально-адаптивної навігації передбачає комплексний підхід, який базується не лише на оптимізації кінематичних параметрів руху, а й на дотриманні міжнародних норм безпеки взаємодії між людиною та автономними системами. Теоретичним підґрунтям для формалізації соціально прийнятної поведінки агента є стандарт ISO 13482:2014 щодо вимог безпеки для сервісних роботів [102].

Документ вимагає від розробників ідентифікувати всі потенційні небезпеки, пов'язані з автономною навігацією та можливим зіткненням, і знизити ризики до прийнятного рівня шляхом впровадження наступних механізмів:

- постійний сенсорний моніторинг робочого простору та дотримання безпечної дистанції між мобільним роботом і людиною;
- обмеження лінійної швидкості та механічних зусиль приводу для запобігання травмуванню у разі неминучої колізії;
- алгоритмічне запобігання соціально небезпечним маневрам, таким як рух у сліпих зонах людини або раптове фронтальне зближення;

- забезпечення високої відмовостійкості систем керування для гарантованої ідентифікації динамічних перешкод у режимі реального часу.

Ефективність соціально-адаптивної навігації безпосередньо залежить від здатності автономної системи мінімізувати когнітивне навантаження на оточуючих людей [103]. Оскільки людський мозок безперервно здійснює прогноз динаміки навколишніх об'єктів для забезпечення власної безпеки, будь-які непередбачувані або хаотичні маневри робота сприймаються як потенційна загроза і підвищує рівень стресу суб'єкта. У цьому контексті критичне значення має не лише фізична дистанція, а й напрямок зближення. Найбільш дискомфортним сценарієм визначено рух робота безпосередньо назустріч людині. Така траєкторія руху робота змушує людину примусово змінювати напрямок руху або зупинятися. Отже, забезпечення передбачуваності та безперервності руху є фундаментальною умовою для психологічно комфортної взаємодії в системі «людина-робот».

Для практичної реалізації цих вимог у розроблену модель навігації інтегровано просторові критерії, які базуються на теорії проксемічних зон Е. Холла. Модель диференціює простір навколо людини на функціональні сегменти, кожному з яких відповідає певний режим роботи АМР. Такий підхід дозволив формалізувати критерії соціальної адаптивності, обравши ключовими параметрами оптимізації дотримання проксемічних дистанцій.

Для формалізації задачі навігації у динамічному середовищі застосовано апарат марковських процесів прийняття рішень (MDP).

Вектор стану S_t у момент часу t містить сенсорні дані, цільові та кінематичні параметри, а також інформацію про динамічні об'єкти середовища [104]

$$S_t = [d_g, \varphi_g, v_r, \omega_r, L_t, H_t], \quad (2.21)$$

де d_g , φ_g – нормалізовані відстань та кут до цілі у локальній системі координат АМР; v_r, ω_r – нормалізовані лінійна та кутова швидкість робота; $L_t \in R^N$ – нормалізований вектор N значень лазерного далекоміра; H_t – вектор станів динамічних агентів, M – кількість динамічних агентів у середовищі.

Вектор станів динамічних агентів складається з індикатора видимості, відносних позицій та швидкостей у локальній системі координат робота $(v_{vis}, x_i, y_i, v_{xi}, v_{yi})$, де координати та швидкості нормалізуються.

Бінарний просторовий індикатор видимості $v_{vis} \in \{0, 1\}$ є дискретною змінною стану, інтегрованою для формалізації наявності прямої лінії спостереження між сенсорною підсистемою автономного мобільного робота та динамічним агентом. Додавання v_{vis} до вектора спостережень забезпечує політику релевантним контекстом щодо фізичної доступності об'єктів. Застосування даного механізму дозволяє алгоритму глибинного навчання з підкріпленням диференціювати потенційно небезпечних агентів від тих, взаємодія з якими надійно ізольована архітектурними об'єктами, виключаючи тим самим генерацію хибних штрафів за порушення соціальної дистанції крізь стіни.

Простір дій A є неперервним $A \subset R^2$. Робот генерує дії, які інтерпретуються як керуючі сигнали для контролера руху $a_t = [v_t, \omega_t]$, де v_t – лінійна швидкість, ω_t – кутова швидкість.

Функція імовірностей переходів $P(s_{t+1} | s_t, a_t)$ у запропонованій моделі визначає закон зміни стану середовища та АМР під впливом обраної дії a_t . У контексті навігації мобільного робота функція $P(s_{t+1} | s_t, a_t)$ визначається кінематичною моделлю платформи, яка встановлює зв'язок між керуючими сигналами та зміною просторових координат робота.

Для реалізації дослідження обрано мобільну платформу з диференціальним приводом [9]. Стан робота у глобальній системі координат описується вектором $q = [x, y, \theta]^T$ де (x, y) – декартові координати геометричного центра платформи, а θ – кут орієнтації відносно осі абсцис.

Перехід системи до наступного стану s_{t+1} здійснюється згідно з рівняннями кінематики в дискретному часі

$$\begin{cases} x_{t+1} = x_t + v_t \cos \theta_t \Delta t, \\ y_{t+1} = y_t + v_t \sin \theta_t \Delta t, \\ \theta_{t+1} = \theta_t + \omega_t \Delta t, \end{cases} \quad (2.22)$$

де v_t та ω_t – лінійна та кутова швидкості, Δt – крок дискретизації часового інтервалу [105].

Врахування кінематичних обмежень у функції переходів дозволяє моделі P коректно відображати фізичні можливості робота, зокрема неможливість миттєвого переміщення у довільному напрямку (неголономність системи). Це змушує політику враховувати геометрію руху при оптимізації функції винагороди. Таким чином, функція імовірностей переходів $P(s_{t+1} | s_t, a_t)$ інтегрує фізичні параметри об'єкта керування в процес інтелектуального прийняття рішень, гарантуючи релевантність апроксимації параметрів динамічного середовища реальним умовам експлуатації АМР.

Центральним елементом запропонованої моделі є функція винагороди R , яка є критерієм оптимізації та спрямовує навчання робота на досягнення цільових показників. Запропоновано комплексну функцію винагороди. Функція винагороди складається з компонентів, які відповідають за окремий аспект поведінки агента:

$$R_t = R_{term,t} + R_{efficiency,t} + R_{social,t}. \quad (2.23)$$

Термінальна складова винагороди $R_{term,t}$ призначена для глобальної оцінки результативності поведінки АМР по завершенні симуляційного епізоду. На відміну від щільних метрик, орієнтованих на покрокову оптимізацію траєкторії руху, дана складова генерує розріджені сигнали виключно в момент переходу системи до одного з кінцевих станів.

$R_{term,t}$ є алгебраїчною сумою трьох компонентів

$$R_{term,t} = R_{coll,t} + R_{goal,t} + R_{timeout,t} \quad (2.24)$$

Компонент винагороди $R_{coll,t}$ спрямований на гарантування фізичної безпеки агента шляхом уникнення контакту із статичними або динамічними перешкодами. Безпека є пріоритетною вимогою в задачах навігації.

$$R_{coll,t} = \begin{cases} -\lambda_{coll}, & \text{якщо } d_{obs} \leq d_{coll} \\ 0, & \text{якщо } d_{obs} > d_{coll} \end{cases}, \quad (2.25)$$

де R_{coll} – штраф за зіткнення; λ_{coll} – константа термінального штрафу за зіткнення, $\lambda_{coll} > 0$.

Якщо відстань до перешкоди d_{obs} стає меншою або рівною за критичний поріг d_{coll} , АМР отримує значний негативний сигнал R_{coll} . Зіткнення визначається як термінальний стан і епізод навчання завершується.

Компонент цільової винагороди $R_{goal,t}$ є стимулом у марковському процесі прийняття рішень, орієнтованим на успішне виконання глобального навігаційного завдання. Дана складова належить до класу розріджених винагород. Її генерація відбувається одноразово, виключно при переході системи до позитивного термінального стану.

Математично нарахування цільового заохочення формалізується як кусково-задана функція, яка залежна від факту входження геометричного центра мобільного робота у визначену цільову область простору

$$R_{goal,t} = \begin{cases} \lambda_{goal}, & \text{якщо } d_{goal} \leq d_{target}, \\ 0, & \text{якщо } d_{goal} > d_{target}, \end{cases} \quad (2.26)$$

де λ_{goal} – скалярна константа термінального позитивного підкріплення; d_{goal} – поточна відстань між центрами мобільного робота та цільової точки; d_{target} – радіус досягнення цілі.

Призначення додатного підкріплення наприкінці успішного епізоду відіграє ключову роль у процесі оптимізації політики DRL. Запропонований механізм дозволяє агенту компенсувати накопичені під час руху від’ємні значення штрафів. Відповідно, алгоритм навчання з підкріпленням змушений формувати поведінку, яка спрямовану на максимізацію загальної кумулятивної винагороди шляхом знаходження коротшого і безпечнішого маршруту до заданої координати, відкидаючи стратегії пасивного блукання.

Для запобігання неефективних дій робота в локальних мінімумах встановлено обмеження кроків на тривалість епізоду T_{max} . У разі перевищення максимально допустимої кількості ітерацій епізод примусово завершується із нарахуванням термінального штрафу $R_{timeout}$

$$R_{timeout} = \begin{cases} -\lambda_{timeout}, & \text{якщо } t > T_{max}, \\ 0, & \text{якщо } t \leq T_{max}, \end{cases} \quad (2.27)$$

де $\lambda_{timeout}$ – скалярна константа.

Даний компонент відсікає нерезультативні траєкторії, змушуючи нейронну мережу шукати ефективні шляхи досягнення цілі.

Компонент винагороди $R_{efficiency,t}$ відповідає за ефективність руху робота, мотивуючи його досягати цільової точки за оптимальний час

$$R_{efficiency} = R_{progress} + R_{time}, \text{ якщо } d_{goal} > d_{target}, \quad (2.28)$$

де $R_{progress}$ – цільна винагорода за прогрес; R_{time} – штраф за час.

Складова винагороди $R_{progress}$ забезпечує постійний зворотний зв'язок. Вона є позитивною, якщо робот скоротив відстань до цілі порівняно з попереднім кроком, і негативною, якщо віддалився. $R_{progress}$ розраховується за формулою

$$R_{progress} = k_g \cdot (d_{g,t-1} - d_{g,t}), \quad (2.29)$$

де $d_{g,t}$ – відстань до цілі в момент часу t ; k_g – ваговий коефіцієнт.

Для забезпечення оптимізації часу проходження маршруту та уникнення нецільових переміщень, структуру функції винагороди доповнено від'ємним компонентом за часовий крок

$$R_{time} = -\lambda_{time}, \quad (2.30)$$

де λ_{time} – константа.

Наявність від'ємного значення R_{time} сприяє максимізації цільової функції шляхом зменшення загальної кількості кроків T до моменту досягнення цілі.

Компонент винагороди R_{social} є частиною моделі, яка трансформує класичну задачу навігації у соціально-адаптивну і формує комфортну для людей поведінку AMP

$$R_{social} = \begin{cases} R_{prox_int}, & \text{якщо } d_h < d_{int}, \\ R_{prox_pers}, & \text{якщо } d_{int} < d_h < d_{pers}. \end{cases} \quad (2.31)$$

де R_{social} – штраф за порушення проксемічних зон. Логіка проксеміки розділена на зони з різним нарахуванням штрафу.

Штраф R_{prox_int} – це штраф за знаходження в інтимному просторі людини з радіусом d_{int} , який розраховується за умови $d_h < d_{int}$

$$R_{prox_int} = -k_{int} \left(\frac{d_{int}-d_h}{d_{int}-k_{coll}} \right)^2 - C_{bias}, \quad (2.32)$$

де d_h – поточна відстань між геометричним центром автономної мобільної платформи та центром найближчого динамічного агента (людини), розрахована на поточному кроці дискретного часу t ; d_{int} – радіус інтимного проксемічного простору людини; k_{int} , k_{coll} – коефіцієнти.

Введення адитивної константи C_{bias} у структуру штрафу за порушення інтимного простору обумовлено необхідністю створення критичного порогу від’ємного підкріплення. Це забезпечує негайну реакцію алгоритму навчання на факт перетину межі простору. Такий підхід дозволяє нівелювати позитивний вплив компоненти винагороди за прогрес $R_{progress}$, змушуючи агента надавати пріоритет стратегіям обходу людини, навіть за умови збільшення загального шляху до цілі.

R_{prox_pers} – штраф за перебування в особистому просторі людини

$$R_{prox_pers} = -k_{pers} \left(\frac{d_{pers}-d_h}{d_{pers}-d_{int}} \right), \quad (2.33)$$

де d_{pers} – радіус персонального проксемічного простору людини; k_{pers} – коефіцієнт масштабування.

Випереджальна стратегія запобігання конфліктним ситуаціям стимулює робота здійснювати маневри зміни курсу на ранніх етапах зближення. Такий підхід підвищує передбачуваність руху в динамічних сценаріях.

Таким чином, розроблена комплексна функція винагороди є багатокритеріальною структурою, яка враховує фізичну безпеку, часову ефективність та дотримання соціальних норм. Інтеграція штрафів дозволяє трансформувати технічну задачу обходу перешкод у модель людино-орієнтованої поведінки. Стратегія керування, яка формується в процесі навчання, безпосередньо залежить від інтерпретації агентом накопиченої винагороди і визначається налаштуваннями часових параметрів MDP.

Гіперпараметр γ процесу навчання визначає горизонт планування агента: низьке значення γ фокусує робота на отриманні миттєвих винагород і нехтуванні довгостроковою безпекою і, навпаки, значення γ , близьке до 1, дозволяє враховувати відтерміновані наслідки кожної дії АМР. Тобто коефіцієнт γ дозволяє регулювати баланс між миттєвою реакцією на динамічні перешкоди та стратегічним плануванням соціально прийнятної траєкторії.

Таким чином, розроблена модель соціально-адаптивної навігації містить комплексну функцію винагороди, яка поєднує вимоги фізичної безпеки, навігаційної ефективності та соціальної прийнятності. Модель створює необхідне теоретичне підґрунтя для подальшої програмної реалізації системи навігації та навчання агента з використанням алгоритмів глибокого навчання з підкріпленням.

2.5. Curriculum Learning як метод оптимізації навчання автономних агентів

У сучасній автономній робототехніці застосування методів навчання з підкріпленням зумовлює виникнення викликів при вирішенні навігаційних задач у високодинамічних середовищах. Апріорне навчання АМР у складних сценаріях часто виявляється неефективним і призводить до низької швидкості збіжності алгоритмів, потрапляння у локальні мінімуми або нездатності знайти оптимальну стратегію поведінки.

Ефективним шляхом вирішення окреслених проблем є застосування методології Curriculum Learning (CL), яка дозволяє оптимізувати навчальний процес [106]. Сутність Curriculum Learning полягає в ієрархічній організації навчального процесу, який імітує когнітивний розвиток людини: від простих прикладів до складних концепцій.

CL передбачає декомпозицію глобальної задачі на послідовність підзадач (curriculum), складність яких зростає ітеративно. Такий підхід дозволяє агенту на початкових етапах сформувати базову поведінку (наприклад, рух без перешкод), а

згодом – здійснювати узагальнення отриманого досвіду для адаптації до складніших умов [107]. У контексті навігаційних систем це забезпечує:

- прискорення процесу навчання;
- підвищення стабільності градієнтного спуску;
- підвищення здатності узагальнення моделі при роботі в раніше невідомих конфігураціях середовища.

Curriculum Learning визначається як стратегія оптимізації моделі машинного навчання, яка базується на динамічному керуванні розподілом навчальних даних [108]. Навчальний процес структурується як впорядкована послідовність критеріїв-завдань із поступовим підвищенням ентропії. Відповідно, розподіл навчальних вибірок набуває вищого рівня складності та варіативності, знижуючи початкову передбачуваність середовища для моделі [109]

$$C = \langle Q_1, \dots, Q_t, \dots, Q_T \rangle. \quad (2.34)$$

Формально, кожен проміжний критерій Q_t на кроці t утворюється шляхом зважування цільового розподілу даних $P(z)$

$$Q_t(z) \propto W_t(z)P(z), \quad (2.35)$$

де $z \in D$ – приклад із навчальної вибірки.

Послідовність завдань повинна задовольняти трьом ключовим умовам:

1. Динаміка навчального процесу повинна характеризуватися монотонним зростанням ентропії розподілів $H(Q_t)$, яка забезпечує послідовне підвищення рівня невизначеності та варіативності вхідних даних для нейронної мережі

$$H(Q_t) < H(Q_{t+1}). \quad (2.36)$$

2. Ймовірність (вага) включення будь-якого конкретного складного прикладу у навчальну вибірку повинна збільшуватися в процесі навчання

$$W_t(z) \leq W_{t+1}(z), \forall z \in D. \quad (2.37)$$

3. На термінальній стадії навчання необхідною умовою є тотожність розподілу навчальних завдань Q_T вихідному розподілу $P(z)$

$(Q_T(z) = P(z))$, що гарантує здатність моделі оперувати в повному просторі станів.

Фундаментальна відмінність CL від інших парадигм навчання полягає у керуванні розподілами навчальних (Н) та тестових (Т) даних. Порівняльний аналіз основних підходів наведено на рис. 2.2, де Н – навчальна вибірка, Т – тестова вибірка, H_j – специфіковані дані для окремих підзадач, $H^{(i)}$ – актуальний розподіл даних на i -му кроці, $Y^{(i)}$ – поточний стан інтелектуального агента [110].

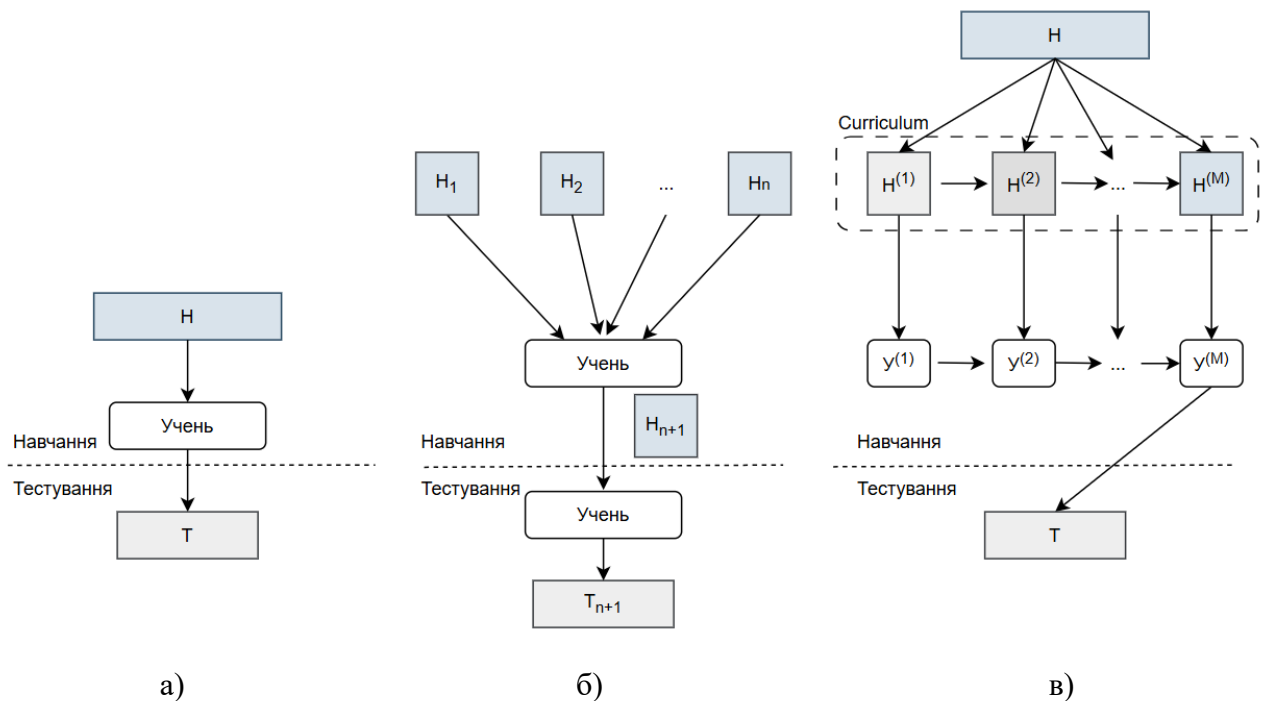


Рис. 2. 2. Схеми навчання: а) Machine Learning, б) Transfer Learning, в) Curriculum Learning.

Аналіз представлених схем дозволяє виділити характерні особливості формування інформаційного простору навчання для кожної парадигми. Зокрема, традиційне машинне навчання (рис. 2.2, а) базується на припущенні про стаціонарність середовища. Воно передбачає, що навчальна та тестова вибірки генеруються з одного і того ж розподілу ($P_{train} \approx P_{test}$). Такий підхід є ефективним для вузькоспеціалізованих задач, проте демонструє обмеженість у динамічних системах, де агент оперує в умовах високої варіативності вхідних сигналів.

Концепція Transfer Learning (рис. 2.2, б) репрезентує гнучкий підхід, побудований на використанні базових знань із вихідного середовища для адаптації

агента до умов цільового завдання. Навіть за наявності розбіжностей у розподілах тренувальних вибірок, Transfer Learning дозволяє переносити вже наявні навички навігації, мінімізуючи обчислювальні витрати та прискорюючи загальний процес навчання.

Стратегія Curriculum Learning (рис. 2.2, в) використовує динамічний розподіл $H^{(i)}$, який еволюціонує в часі. Це дозволяє учневі $U^{(i)}$ на кожному кроці i послідовно адаптуватися до зростаючого ускладнення умов, поки характеристики навчального розподілу $H^{(i)}$ не стануть тотожними цільовому середовищу. Такий підхід забезпечує стабільність та мінімізує ризик застрягання агента у локальних мінімумах на ранніх етапах навчання.

У контексті навігації мобільних роботів методологія Curriculum Learning реалізується у вигляді ітераційної модифікації середовища або декомпозиції функції винагороди [111].

Ітераційна модифікація середовища ґрунтується на генерації впорядкованої послідовності тренувальних просторів $\{E_1, E_2, \dots, E_n\}$. У цій структурі E_1 – елементарний простір із мінімальною кількістю перешкод, тоді як E_n відповідає цільовому середовищу. Рівень складності в такій послідовності може здійснюватись шляхом ітераційної модифікації навчальних умов середовища:

1. Модифікація просторових та топологічних характеристик, яка полягає в ітераційному ускладненні конфігурації середовища шляхом підвищення щільності статичних перешкод або зміни архітектури приміщення. Зокрема, передбачено поетапний перехід від відкритих просторів до структурно складних зон, таких як вузькі проходи та зони обмеженої прохідності. Такий підхід дозволяє моделі вивчати ефективні стратегії локального маневрування та орієнтації в обмеженому просторі [112].
2. Регулювання динаміки та стохастичності сцени здійснюється шляхом введення в середовище мобільних агентів з поступовим збільшенням їхньої кількості та варіюванням векторів швидкості від лінійних до непередбачуваних траєкторій. Додатково для підвищення робастності

моделі до реальних умов експлуатації у вихідні дані сенсорів впроваджується адитивний стохастичний шум [107].

3. Зміна параметрів генерації стартових і цільових позицій, яка відбувається за рахунок поступового збільшення відстані між роботом та заданою ціллю. Такий підхід нівелює недоліки розрідженої винагороди на старті навчання, стимулюючи систему до довгострокового планування траєкторій у межах простору станів [113].

Починаючи навчання у спрощеному середовищі, агент швидше засвоює базові кінематичні навички та алгоритми уникнення статичних перешкод. Це мінімізує час дослідження та дозволяє поступово адаптувати політику до нових факторів і забезпечити вищу підсумкову успішність.

Основними недоліками ітераційної модифікації середовища є складність автоматизації переходів між етапами та ризик «забування» базових навичок при надмірному фокусуванні на нових сценаріях. Традиційно перехід між рівнями складності здійснюється вручну або за жорсткими порогами (кількість кроків, відсоток успішного досягнення цілі) і не завжди враховує реальну стабільність засвоєної політики.

Успішна збіжність моделі в умовах багатовимірного простору станів визначається не лише складністю середовища, а й інформативністю сигналу підкріплення, який формує траєкторію навчання агента. Тому важливим аспектом є застосування методу формування функції винагороди (Reward Shaping), який інтегрується в загальну структуру Curriculum Learning. Методологія передбачає модифікацію сигналу підкріплення для створення допоміжних стимулів пошуку оптимальної стратегії π . Формування функції винагороди є динамічним регулятором складності задачі [109].

У межах структури Curriculum Learning можливе застосування двох стратегій модифікації функції винагороди: поступове спрощення підкріплення до розрідженої форми та поступове нарощування складності критеріїв оптимізації.

Однією із основних проблем навчання AMP у складних середовищах є проблема розрідженості винагороди, тобто агент при виконанні тисячі дій, може не

отримати жодного зворотного підкріплення. Підхід формування винагороди вирішує цю проблему шляхом трансформації вихідної функції винагороди $R(s, a, s')$ у модифіковану функцію $R'(s, a, s')$, де

$$R'(s, a, s') = R(s, a, s') + F(s, a, s'), \quad (2.38)$$

де $F(s, a, s')$ – функція формування, яка визначається як різниця потенціалів станів

$$F(s, a, s') = \gamma\Phi(s') - \Phi(s), \quad (2.39)$$

де Φ – потенціальна функція, яка відображає наближення до мети;

γ – коефіцієнт дисконтування.

Використання вказаної форми винагороди гарантує, що модифікація винагороди не змінить множину оптимальних стратегій, але прискорить швидкість їх знаходження [114].

Підхід поступового спрощення підкріплення до розрідженої форми реалізується через перехід від щільної винагороди до розрідженої:

1. На етапі ініціалізації знань функція винагороди містить проміжні стимули, які створюють стійкий градієнт у напрямку цільового стану. Це дозволяє мінімізувати час випадкового дослідження простору станів та пришвидшити збіжність моделі.
2. На етапі стабілізації здійснюється ітераційне зниження впливу допоміжних стимулів відповідно до прогресу навчання. Такий підхід є умовою коректного формування стратегії агента. У результаті забезпечується фокусування алгоритму на досягненні фінальної мети дослідження, а не на отриманні локальних винагород. Це дозволяє уникнути явища «зламу винагороди», коли агент оптимізує стратегію для максимізації проміжних компонентів функції формування, ігноруючи при цьому досягнення глобальної цілі навігації.
3. На стадії кінцевої адаптації агент переходить до навчання на оригінальній, розрідженій функції винагороди. Під час виконання завдань система використовує попередньо сформовану поведінку, яка забезпечує

стабільну роботу інтелектуального агента за умов низької інформативності сигналу підкріплення.

Підхід на основі поступового збільшення складності функції винагороди формалізує винагороду як багаторівневу математичну модель, складність якої збільшується синхронно з процесом оптимізації політики керування.

На початкових етапах навчання використовується мінімальний набір критеріїв, які забезпечують формування базової цільової поведінки, тоді як наступні етапи передбачають інтеграцію додаткових обмежень. Кожна наступна функція включає нові компоненти оптимізації, які деталізують вимоги до поведінки агента

$$R_1(s, a) \subset R_2(s, a) \subset \dots \subset R_n(s, a). \quad (2.40)$$

Поетапне ускладнення функції винагороди забезпечує зменшення дисперсії оцінки функції цінності на ранніх стадіях навчання та запобігає виникненню конфліктів між численними компонентами підкріплення. Введення складних критеріїв після формування базових навичок дозволяє уникнути перевантаження оптимізаційного процесу та зменшує можливість виникнення локальних мінімумів, зумовлених одночасною оптимізацією взаємоконфліктних показників [115].

У завданнях соціально-орієнтованої навігації АМР стратегія поступового ускладнення функції винагороди дозволяє формувати поведінку АМР відповідно до ієрархії цілей – від забезпечення фізичної безпеки до досягнення соціальної прийнятності та комфортності взаємодії у динамічному середовищі.

Основними перевагами інтеграції методів формування винагороди у структуру CL є прискорення збіжності алгоритму та вирішення проблеми розріджених сигналів підкріплення. Використання проміжних стимулів на початкових етапах створює стійкий градієнт навчання, що дозволяє автономному агенту ефективно досліджувати простір станів і уникати локальних оптимумів. Крім того, поетапне формування винагороди забезпечує стабільність навчання, оскільки дозволяє агенту поступово адаптувати свою політику до нових компонентів.

Водночас даний підхід має недоліки, серед яких це ризик виникнення явища зламу винагороди, коли агент знаходить спосіб максимізувати допоміжні стимули, ігноруючи при цьому глобальну мету дослідження. Також недоліком є складність математичної формалізації та налаштування вагових коефіцієнтів.

2.6. Метод навчання навігаційної політики на основі глибинного навчання з підкріпленням та Curriculum Learning

Для подолання проблеми повільної збіжності алгоритмів DRL у складних соціальних середовищах, запропоновано метод навчання навігаційної політики на основі глибинного навчання з підкріпленням та Curriculum Learning [110]. Теоретичним підґрунтям методу є модель соціально-адаптивної навігації, яка формалізує просторові зони комфорту людини та критерії безпечної поведінки АМР.

Метод забезпечує формування адаптивних поведінкових патернів автономного мобільного робота, мінімізує ризик збіжності до локальних мінімумів на початкових етапах навчання і дозволяє синхронізовано реалізувати дві стратегії поступового ускладнення завдань:

- ітераційної модифікації середовища, яка передбачає контрольоване підвищення ентропії простору станів шляхом поетапного введення динамічних агентів та варіативності сценаріїв;
- поетапного формування функції винагороди, яке базується на поступовій інтеграції соціально-орієнтованих критеріїв оптимізації в структуру підкріплення.

Таким чином, у межах запропонованого підходу процес навчання організовується як координована зміна як параметрів середовища, так і структури функції винагороди, і забезпечується узгоджене зростання складності задачі [116].

Базовим алгоритмом глибинного навчання з підкріпленням, на основі якого формується навігаційна політика в межах розробленого методу, обрано Proximal Policy Optimization (PPO). Вибір цієї архітектури зумовлений її здатністю

забезпечувати високу стабільність оновлення політики та ефективну роботу в умовах безперервних просторів станів і дій.

Навчальний процес організовано через послідовність чотирьох етапів, кожен із яких відповідає визначеному рівню складності задачі. Перехід між рівнями здійснюється автоматизовано на основі встановлених критеріїв та відбувається виключно за умови досягнення стабільних показників продуктивності політики на поточному етапі. Такий механізм забезпечує адаптивне керування складністю навчання та відповідає принципам парадигми Curriculum Learning, відповідно до якої ускладнення задачі синхронізується з рівнем сформованості навичок агента.

На першому етапі навчання процедура генерації стартової та цільової позицій мобільного робота реалізується безпосередньо в межах статичного середовища. АМР опановує базові кінематичні стратегії та принципи навігації у статичному середовищі. Основною метою є формування здатності до цілеспрямованого руху, яка передбачає побудову ефективної траєкторії до цілі та уникнення зіткнень із нерухомими перешкодами.

Функція винагороди на цьому етапі орієнтована переважно на мінімізацію відстані до цілі, уникнення зіткнень і обмеження неефективних кінематичних дій.

Функція винагороди визначається як сума двох компонентів:

$$R_t = R_{\text{term},t} + R_{\text{efficiency},t}, \quad (2.41)$$

де $R_{\text{term},t}$ – термінальна винагорода, а $R_{\text{efficiency},t}$ – винагорода за ефективність руху до цілі.

Результатом першого етапу є формування політики, здатної забезпечувати досягнення цільової позиції у статичному середовищі з мінімальним рівнем зіткнень. Отриманий функціональний базис створює необхідні передумови для подальшого ускладнення середовища.

На другому етапі навчання стартова та цільова позиції мобільного робота генеруються безпосередньо в середовищі із динамічним агентом. При цьому структура функції винагороди залишається незмінною. Це дозволяє підвищити складність задачі виключно за рахунок динаміки середовища. За таких умов АМР

адаптує сформовану політику до рухомих перешкод, враховуючи вимоги фізичної безпеки.

На третьому етапі навчання стратегія Curriculum Learning фокусується на якісній трансформації моделі поведінки АМР через перехід від реактивного уникнення зіткнень до соціальної адаптивності поведінки робота. На цьому етапі здійснюється структурна модифікація функції винагороди шляхом інтеграції компонентів соціально-орієнтованих обмежень.

Функцію винагороди розширено шляхом впровадження соціального компонента $R_{social,t}$.

$$R_t = R_{term,t} + R_{efficiency,t} + R_{social,t}, \quad (2.42)$$

де $R_{social,t}$ відображає дотримання соціально прийнятної поведінки АМР.

На завершальному етапі навчання стратегія Curriculum Learning фокусується на масштабуванні простору та збільшенні чисельності динамічних агентів. Метою даного етапу є формування здатності АМР до узагальнення політики в умовах підвищеної щільності динамічних перешкод і складної соціальної взаємодії.

Блок-схему запропонованого методу наведено на рис. 2. 3.

Автоматизований перехід між етапами навчання реалізовано на основі аналізу часової динаміки показника ефективності політики. Для оцінювання ефективності політики використано епізодичну винагороду:

$$J_e = \sum_{t=0}^{T_e} R_t, \quad (2.43)$$

де T_e – тривалість епізоду.

Аналіз динаміки навчання здійснено у ковзному часовому вікні W_l , визначеному у просторі глобальних кроків середовища довжиною N_l . При цьому статистичні оцінки обчислено за множиною епізодів E_l , які завершилися в межах цього вікна. M_l – кількість елементів у множині E_l .

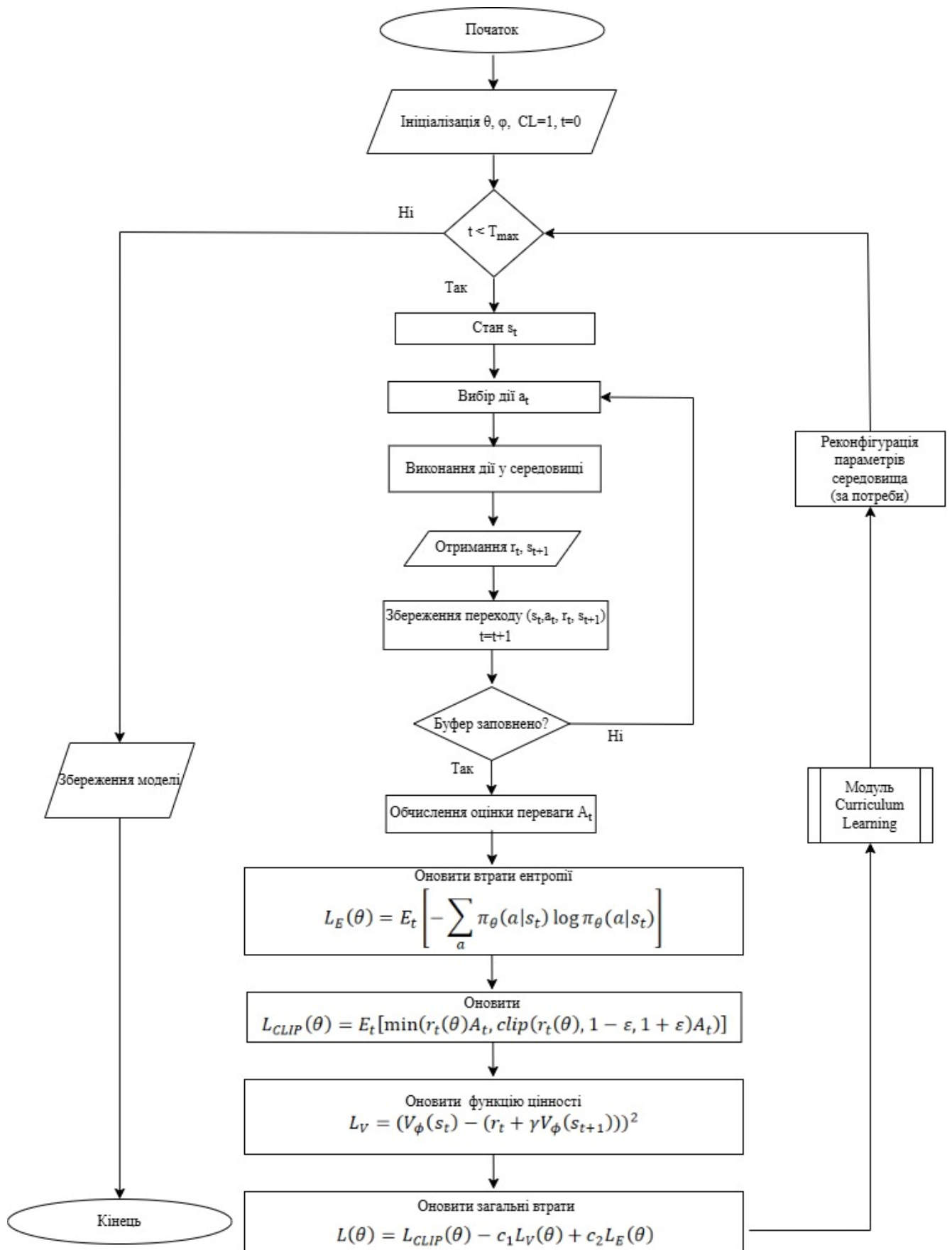


Рис.2.3. Блок-схема методу соціально-адаптивної навігації на основі DRL та Curriculum Learning

Середнє значення епізодичної винагороди визначається як:

$$\bar{J}_l = \frac{1}{M_l} \sum_{e \in E_l} J_e. \quad (2.44)$$

Стандартне відхилення визначається за формулою:

$$\sigma_l = \sqrt{\frac{1}{M_l} \sum_{e \in E_l} (J_e - \bar{J}_l)^2}. \quad (2.45)$$

Для оцінки стаціонарності процесу навчання використовується різниця між середніми значеннями у двох послідовних вікнах. Такий підхід дозволяє оцінювати якість керування

$$\Delta \bar{J}_l = |\bar{J}_l^{current} - \bar{J}_l^{prev}|. \quad (2.46)$$

Додатково враховується показник успішності:

$$S_l = \frac{1}{M_l} \sum_{e \in E_l} s_e, \quad (2.47)$$

де $s_e \in \{0,1\}$ – індикатор успішного завершення епізоду.

Алгоритмічну реалізацію переходу між рівнями складності Curriculum Learning представлено на рис. 2.4.

Перехід між етапами $l \in \{1,2,3\}$ на наступний рівень складності здійснюється автоматизовано за умови досягнення стаціонарності, стабільності та успішності процесу навчання. Умовою переходу є одночасне виконання нерівностей:

$$\Delta \bar{J}_l < \varepsilon_l \wedge \sigma_l < \delta_l \wedge S_l > \eta_l, \quad (2.48)$$

де $\varepsilon_l, \delta_l, \eta_l$ – порогові значення для відповідного етапу.

Перевірка умов переходу здійснюється дискретно з фіксованим інтервалом, за умови накопичення достатнього обсягу статистичних даних, зокрема $M_l \geq M_l^{min}$ та наявності двох послідовних часових вікон після початку поточного етапу. У разі виконання критеріїв процес навчання вважається збіжним, після чого ініціюється перехід до наступного рівня складності.

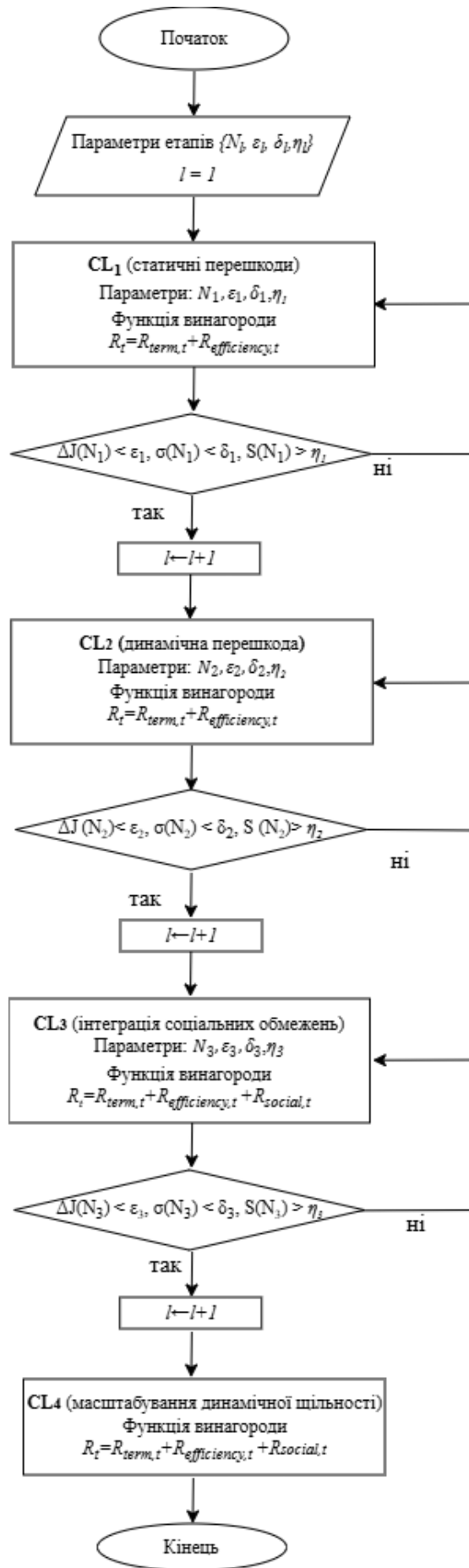


Рис. 2.4. Блок-схема модуля Curriculum Learning

Для кожного етапу навчання l використовуються індивідуальні параметри контролю збіжності $N_l, \varepsilon_l, \delta_l$. Числові значення параметрів контролю збіжності для кожного етапу навчання l встановлюються експериментальним шляхом на етапі налаштування гіперпараметрів моделі. Це дозволяє адаптувати критерій переходу до специфіки відповідного рівня складності та характеристик середовища.

Оскільки поетапна інтеграція динамічних агентів та модифікація функції винагороди призводить до зростання стохастичності середовища та дисперсії сигналу підкріплення, параметри N_l, δ_l та ε_l зростають для кожного наступного рівня складності.

На початкових етапах зі статичними перешкодами сигнал підкріплення є переважно детермінованим, що забезпечує низький рівень дисперсії епізодичної винагороди та сприяє швидкій збіжності політики. Водночас інтеграція динамічних агентів та компонентів функції винагороди на наступних етапах навчання зумовлює суттєве зростання варіативності сигналу підкріплення. Це обумовлено як стохастичною природою поведінки динамічних агентів, так і нелінійністю соціальних штрафів. В таких умовах застосування однакових критеріїв збіжності є недоцільним, оскільки варіації винагороди зумовлені не лише якістю політики, але й внутрішньою невизначеністю середовища.

Збільшення розміру вікна N_l дозволяє підвищити статистичну достовірність оцінки ефективності політики шляхом згладжування випадкових флуктуацій винагороди. Водночас зростання допустимого стандартного відхилення δ_l запобігає блокуванню переходу між етапами в умовах природно високої дисперсії сигналу підкріплення.

Відповідно, збільшення порогового значення ε_l , яке визначає допустиму зміну середньої епізодичної винагороди, виконує функцію регулювання критерію стаціонарності та забезпечує:

- запобігання надмірному перенавчанню політики під конкретні випадкові траєкторії динамічних перешкод у межах поточного рівня;

- підвищення ефективності використання обчислювальних ресурсів шляхом зупинки навчання на етапі досягнення стабільної поведінки, коли подальше зростання епізодичної винагороди є незначним;
- своєчасний перехід до наступного рівня складності.

Реалізація запропонованого методу дозволяє організувати процес навчання як послідовність етапів із ітераційною модифікацією середовища та поетапного ускладнення функції винагороди. Такий підхід забезпечує стабільну збіжність навігаційної політики в умовах високої ентропії простору станів. Результатом реалізації запропонованого методу є оптимізована політика навігації AMP.

Висновки до розділу 2

У другому розділі здійснено теоретико-методологічне обґрунтування основ моделювання руху автономного мобільного робота в динамічному соціальному середовищі. За результатами проведених досліджень сформульовано такі висновки:

Виконано формалізацію проксемічних обмежень на основі теорії зон Е. Холла, що дозволило визначити кількісні критерії соціальної прийнятності маневрів робота. Впровадження математичних моделей інтимного, особистого та соціального простору у алгоритми планування забезпечує формування траєкторій, які мінімізують психологічний дискомфорт оточуючих людей.

Здійснено математичне моделювання процесу навігації. Задачу керування мобільним роботом у неструктурованому середовищі формалізовано як марковський процес прийняття рішень (MDP). Такий підхід дозволив звести задачу навігації до оптимізаційної задачі максимізації очікуваної винагороди, що обґрунтовує застосування методів глибинного навчання з підкріпленням.

Удосконалено модель соціально-адаптивної навігації. Ключовим елементом моделі є запропонована комплексна функція винагороди, яка інтегрує три групи компонентів: оцінку термінальних станів, ефективність виконання завдання та

соціальну прийнятність. Це дозволяє трансформувати неявні соціальні норми у чіткі математичні обмеження для навчання нейронної мережі.

Розроблено метод навчання навігаційної політики на основі глибинного навчання з підкріпленням та Curriculum Learning, який здійснює ієрархічну декомпозицію глобальної задачі на впорядковану послідовність із чотирьох етапів. Метод поєднує стратегії ітераційної модифікації середовища та поетапного ускладнення функції винагороди. Запропоновано автоматизований механізм переходу між етапами навчання на основі аналізу стаціонарності, стабільності та успішності процесу навчання.

Таким чином, у другому розділі створено необхідний теоретико-методологічний базис для програмної реалізації та експериментального дослідження системи соціально-адаптивної навігації.

РОЗДІЛ 3

МЕТОД АДАПТИВНОГО ФОРМУВАННЯ ВИНАГОРОДИ НА ОСНОВІ ПРОГНОЗУ НЕВИЗНАЧЕНОСТІ ДИНАМІЧНОГО СЕРЕДОВИЩА

3.1. Аналіз архітектурних рішень для систем соціальної навігації в умовах невизначеності середовища

Ефективна соціальна навігація автономного мобільного робота в середовищі з присутністю людей вимагає вирішення багатокритеріальної оптимізаційної задачі зі змінними пріоритетами. Складність дослідження обумовлена необхідністю підтримувати баланс між уникненням зіткнень, мінімізацією часу досягнення цілі і довжини шляху та дотриманням соціальних норм.

В динамічному середовищі складність задачі посилюється стохастичною природою людської поведінки. Для розробки адаптивного агента важливим є наявність механізму, здатного оцінювати невизначеність руху людини та динамічно коригувати стратегію поведінки.

Аналіз існуючих методів моделювання стохастичної невизначеності траєкторій дозволяє виділити такі основні підходи, які застосовуються в сучасній робототехніці: генеративно-змагальні мережі (GAN), варіаційні автоенкодера (VAE), рекурентні ймовірнісні моделі LSTM-MDN та моделі на основі механізмів уваги.

Одним із домінуючих підходів є використання генеративних змагальних мереж, зокрема архітектури Social-GAN [65]. Перевагою таких моделей є здатність ефективно вирішувати проблему «усереднення» прогнозів, генеруючи чіткі, соціально прийнятні мультимодальні траєкторії. Проте, для оцінки невизначеності система змушена генерувати значну кількість семплів з подальшим статистичним аналізом отриманої вибірки, що створює надмірне обчислювальне навантаження. Крім того, GAN схильні до проблеми «колапсу мод», коли ігноруються рідкісні, але важливі варіанти поведінки.

Альтернативним імовірнісним підходом є використання умовних варіаційних автоенкодерів (CVAE), наприклад Social-CVAE [117], які описують невизначеність через прихований простір змінних. Порівняно з GAN, ці архітектури демонструють більш стабільний процес навчання та дозволяють явно моделювати розподіл латентних змінних, охоплюючи широкий спектр людських намірів. Водночас, подібно до генеративних мереж, отримання метрики невизначеності U вимагає багаторазового прогону декодера для оцінки дисперсії траєкторій. Цей процес створює додаткову затримку в контурі керування, яка може негативно вплинути на безпеку руху робота в соціальному середовищі.

Методи на основі графів, такі як Social-BiGAT [118], моделюють сцену як граф взаємодій. Завдяки детальному врахуванню топології соціальних зв'язків та впливу оточення, такі моделі забезпечують високу точність прогнозування у щільних натовпах. Проте основною проблемою для їх імплементації на мобільних роботах є висока обчислювальна складність. Залежно від кількості агентів обчислювальне навантаження нелінійно зростає і є критичним фактором для систем з обмеженими ресурсами.

Окремий клас підходів, який поєднує часове моделювання та ймовірнісну інтерпретацію, представлений архітектурами LSTM-MDN (Long Short-Term Memory – Mixture Density Network), запропонованими в роботах на основі ідей К. Бішопа. У таких моделях рекурентна мережа типу LSTM використовується для вилучення часових залежностей з історії руху агента, тоді як вихідний шар MDN параметризує суміш гаусових розподілів, що описують можливі майбутні стани.

Ключовою перевагою підходу LSTM-MDN є можливість отримання параметрів повного умовного розподілу за один прохід прямого поширення сигналу через нейронну мережу [119]. Це дозволяє безпосередньо оцінювати невизначеність прогнозу без необхідності багаторазового семплювання, як у випадку GAN або CVAE. Крім того, модель підтримує мультимодальність, що є критично важливим для опису альтернативних сценаріїв поведінки людини.

Разом з тим, підхід LSTM-MDN має і певні обмеження. Зокрема, якість апроксимації складних розподілів суттєво залежить від кількості компонент

суміші, що може впливати на стабільність навчання. Крім того, на відміну від графових або attention-based моделей, базова LSTM-MDN не враховує явно соціальні взаємодії між агентами і потребує додаткового розширення архітектури для використання в соціальних середовищах.

Таким чином, LSTM-MDN представляє компроміс між обчислювальною ефективністю та якістю моделювання невизначеності, що робить його перспективним для використання в системах реального часу автономної соціальної навігації.

Результати порівняльного аналізу методів моделювання невизначеності за визначеними критеріями зведено у табл. 3.1.

Таблиця 3.1.

Методи моделювання невизначеності траєкторій

Метод	Механізм моделювання невизначеності	Отримання метрики невизначеності (U)	Обчислювальна ефективність	Придатність для RL-винагороди
LSTM-MDN	Суміш гаусових розподілів (GMM)	Пряме обчислення з параметрів суміші	Висока (один прямий прохід)	Висока
Social-GAN	Генеративно-змагальна мережа (GAN)	Непряма оцінка через дисперсію згенерованих семплів	Низька (потрібна генерація великої кількості семплів)	Низька
Social-CVAE	Умовний варіаційний автоенкодер (CVAE)	Непряма оцінка через дисперсію семплів латентного простору	Середня/Низька (залежить від кількості семплів)	Середня
Social-BiGAT	Графова мережа уваги (GAT) у поєднанні з GAN	Непряма оцінка невизначеності	Низька (висока складність графових обчислень)	Низька

3.2. Прогнозування станів динамічних об'єктів з використанням LSTM-MDN

Використання нейронних мереж для прогнозування майбутніх станів динамічних об'єктів базується на мінімізації середньоквадратичної помилки (Mean Squared Error, MSE). Цей підхід генерує єдине усереднене значення для прогнозованої позиції.

У ситуаціях просторової невизначеності, наприклад, коли пішохід має намір обійти перешкоду ліворуч або праворуч з однаковою ймовірністю, застосування MSE призводить до критичних помилок. Мережа намагається мінімізувати відстань до обох можливих варіантів одночасно, генеруючи усереднений прогноз, який може проходити через фізично недопустимі області, наприклад, через перешкоду. Вирішенням цієї проблеми є перехід від детермінованого прогнозування до ймовірнісного моделювання за допомогою архітектури LSTM-MDN, концепцію якої вперше запропонував К. Бішоп [120], а згодом адаптував для послідовних даних А. Грейвс [121].

Першим компонентом гібридної архітектури є рекурентна нейронна мережа з довгою короткочасною пам'яттю (LSTM). Головна функція LSTM полягає у виділенні прихованих просторово-часових залежностей з історії спостережень за рухом об'єктів. На відміну від стандартних рекурентних мереж, LSTM здатна вирішувати проблему зникаючого градієнта завдяки використанню складної системи вентилів, здатних регулювати потік інформації [122].

Процес оновлення стану комірки LSTM на кожному часовому кроці t описується системою матричних рівнянь. На першому етапі вентиль забування (forget gate) f_t визначає частку інформації з попереднього стану комірки C_{t-1} , яку необхідно видалити:

$$f_t = \sigma(W_f(h_{t-1}, x_t) + b_f). \quad (3.1)$$

Далі вхідний вентиль (input gate) i_t визначає, які нові дані будуть збережені у внутрішньому стані, а також створюється вектор кандидатів \tilde{C}_t

$$i_t = \sigma(W_i(h_{t-1}, x_t) + b_i), \quad (3.2)$$

$$\tilde{C}_t = \tanh(W_c(h_{t-1}, x_t) + b_c). \quad (3.3)$$

Оновлення стану комірки C_t відбувається шляхом лінійної комбінації попереднього стану та нових значень-кандидатів

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t. \quad (3.4)$$

На фінальному етапі вихідний вентиль (output gate) o_t обчислює новий прихований стан h_t , який передається на наступний часовий крок та слугує входним сигналом для наступних шарів мережі

$$o_t = \sigma(W_o(h_{t-1}, x_t) + b_o), \quad (3.5)$$

$$h_t = o_t \odot \tanh(C_t), \quad (3.6)$$

де x_t – вектор входних даних на поточному часовому кроці (координати, швидкість об'єкта); h_{t-1} – прихований стан мережі на попередньому кроці; W – матриці вагових коефіцієнтів для відповідних вентилів; b – вектори зсуву (bias); σ – сигмоїдна активаційна функція для масштабування значень у діапазон $[0, 1]$; \tanh – гіперболічний тангенс для масштабування значень у діапазон $[-1, 1]$.

Прихований стан h_t акумулює всю релевантну історію руху пішохода і слугує базою для генерування ймовірнісного прогнозу.

Модуль MDN приймає на вхід прихований стан h_t з рекурентного шару і перетворює його на параметри суміші гаусових розподілів (Gaussian Mixture Model, GMM). Густина ймовірності $p(y_t | x_1, \dots, x_t)$ для майбутньої позиції y_t за умови спостереження попередньої траєкторії моделюється як зважена сума K двовимірних нормальних розподілів

$$p(y_t | h_t) = \sum_{k=1}^K \pi_k(h_t) N(y_t | \mu_k(h_t), \Sigma_k(h_t)). \quad (3.7)$$

Функція двовимірного нормального розподілу N розраховується за формулою

$$N(y_t|\mu, \Sigma) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{\left(\frac{Z}{2(1-\rho^2)}\right)}. \quad (3.8)$$

Внутрішня змінна Z визначається як

$$Z = \frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} - \frac{2\rho(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y}, \quad (3.9)$$

де π_k – ваговий коефіцієнт k -ї компоненти (ймовірність обрання конкретної гіпотези руху); μ_x, μ_y – математичні сподівання координат x та y (центри розподілу); σ_x, σ_y – стандартні відхилення (рівень просторової невизначеності); ρ – коефіцієнт кореляції між координатами x та y .

Для забезпечення математичної коректності прогнозованих параметрів застосовуються спеціальні функції активації на вихідному шарі MDN. Вагові коефіцієнти π_k повинні бути додатними, а їх сума має дорівнювати одиниці, тому використовується функція Softmax

$$\pi_k = \frac{e^{z_{\pi,k}}}{\sum_{j=1}^K e^{z_{\pi,j}}}. \quad (3.10)$$

Стандартні відхилення σ за визначенням є строго додатними величинами, для їх обчислення застосовується експоненційна функція:

$$\sigma_{kx} = e^{z_{\sigma,kx}}. \quad (3.11)$$

Коефіцієнт кореляції ρ лежить у межах $[-1, 1]$, відповідно використовується гіперболічний тангенс

$$\rho_k = \tanh(z_{\rho,k}). \quad (3.12)$$

Навчання гібридної моделі здійснюється за допомогою методу зворотного поширення помилки в часі (Backpropagation Through Time, BPTT). Замість стандартної MSE, цільова функція втрат L (Loss function) формулюється як від'ємна логарифмічна функція правдоподібності (Negative Log-Likelihood, NLL). Мережа мінімізує цю функцію, намагаючись максимізувати ймовірність правильного передбачення реальної траєкторії з навчального набору даних:

$$L(W) = - \sum_{t=1}^T \ln \left(\sum_{k=1}^K \pi_k N(y_t | \mu_k, \Sigma_k) \right). \quad (3.13)$$

Мінімізація NLL дозволяє моделі не лише точно позиціонувати центри гаусових розподілів μ , але й коректно оцінювати дисперсію σ . У разі впевненості моделі у своєму прогнозі дисперсія зменшується, генеруючи вузький пік імовірності. При високій невизначеності дисперсія збільшується, відображаючи широкий спектр можливих позицій об'єкта.

Структурну блок-схему LSTM-MDN наведено на рис. 3.1.



Рис.3.1. Блок-схема LSTM-MDN

У сучасній робототехніці розглянута архітектура є фундаментом для систем соціально-прийнятної навігації. Вагомим кроком у цьому напрямі стала розробка архітектури Social LSTM дослідниками зі Стенфордського університету [62]. Вони запропонували концепцію «соціального пулінгу», яка дозволяє об'єднувати приховані стани h_t рекурентних мереж усіх пішоходів, розташованих у певному радіусі один від одного. Передача таких агрегованих даних на шар MDN дає змогу моделі розуміти правила соціальної взаємодії – уникання зіткнень між самими людьми, рух у потоці та реакцію на наближення автономного робота.

Застосування вихідних розподілів MDN відіграє важливу роль у плануванні безпечних маневрів. Навігаційна система робота розглядає кожну з K компонент гаусової суміші як окрему гіпотезу щодо подальшого шляху пішохода. Автономний агент генерує локальний маршрут, мінімізуючи ймовірність перетину власної траєкторії з еліпсами невизначеності кожної гіпотези, зваженими на їхню ймовірність π_k .

Незважаючи на значні переваги, архітектура LSTM-MDN має певні технічні обмеження. Дослідники О. Макансі та співавтори зазначають проблему колапсу мод, коли мережа ігнорує рідкісні, але можливі варіанти руху, віддаючи всю вагу π_k лише одному домінуючому напрямку [9]. Вирішення цієї проблеми відбувається шляхом додавання регуляризаційних складових до функції втрат або застосуванням методів вибірки. Водночас гібридна модель LSTM-MDN залишається ефективним та обчислювально доцільним підходом для застосування у вбудованих системах автономних роботів з обмеженими обчислювальними ресурсами.

3.3. Модель мультимодального прогнозування станів динамічних об'єктів

Одним із завдань моделювання динамічного середовища АМР є врахування стохастичної природи людської поведінки, яка за своєю суттю є мультимодальною.

З поточного розташування людини існує не одна, а множина вірогідних майбутніх траєкторій, реалізація яких залежить від латентних намірів людини.

Вхідними даними для моделі є послідовність станів агентів, які отримують протягом фіксованого часового вікна історії $T_{history}$. Вектор стану S для кожного агента включає координати (x, y) та швидкості (v_x, v_y) .

Нехай X – вхідна послідовність спостережень:

$$X = \{s_{t-T_{history}}, \dots, s_t\}, \quad (3.14)$$

де s_t – стан агента в момент часу t .

Вихідний шар мережі прогнозує параметри суміші K двовимірних нормальних розподілів для кожного часового кроку τ у майбутньому горизонті прогнозування T_p .

Навчання мережі здійснюється шляхом мінімізації від’ємної логарифмічної функції правдоподібності NLL. Мережа апроксимує мультимодальність розподілу траєкторій агентів-людей. Цільова функція втрат визначається як:

$$L_{NLL} = -\frac{1}{T} \sum_{t=1}^T \ln \left(\sum_{k=1}^K \pi_{k,t} N(y_{target,t} | \mu_{k,t}, \sigma_{k,t}, \rho_{k,t}) \right), \quad (3.15)$$

де K – кількість компонентів суміші; y_{target} – поточна позиція динамічного агента з навчальної вибірки; $\mu_{k,t}$ – математичне сподівання, яке представляє найбільш імовірні координати (x, y) центру k -ї траєкторії; $\sigma_{k,t}$ – середньоквадратичне відхилення, чим більше $\sigma_{k,t}$, тим менш впевнена модель у точному положенні об’єкта; $\rho_{k,t}$ – коефіцієнт кореляції між координатами x та y ; $N(.)$ – функція густини двовимірного нормального розподілу, яка оцінює, наскільки близько реальна позиція y_{target} знаходиться до прогнозованого параметричного розподілу; $\pi_{k,t}$ – ваговий коефіцієнт k -го компонента суміші в момент часу t . Він визначає відносну ймовірність (пріоритет) конкретної траєкторії серед інших можливих варіантів.

$$\sum_{k=1}^K \pi_k = 1. \quad (3.16)$$

Особливістю розробленої математичної моделі є можливість розрахунку метрики невизначеності середовища U в реальному часі без необхідності ресурсоемного семпсування.

Показник невизначеності U_t для N_{max} агентів на горизонті прогнозування T_p розраховується за формулою

$$U_t = \max_n \left(e^{\left(-\frac{d_n^2}{2R_{vis}^2} \right)} \frac{1}{T_p} \sum_{\tau=1}^{T_p} \sum_{k=1}^K \pi_{n,k,\tau} \sqrt{\sigma_{x,k,\tau}^2 + \sigma_{y,n,k,\tau}^2} \right), \quad (3.17)$$

де індекс n відповідає n -му динамічному агенту; d_n – відстань від робота до n -го динамічного агента; R_{vis} – радіус межі сенсорної видимості AMP.

Отримане значення U_t слугує індикатором глобальної невизначеності сцени. Високі значення U_t свідчать про наявність розбіжних гіпотез руху (наприклад, на перехрестях), тоді як низькі значення вказують на високу впевненість моделі у прогнозі.

Структурна схема модуля ймовірнісного прогнозування представлена на рис. 3. 2.

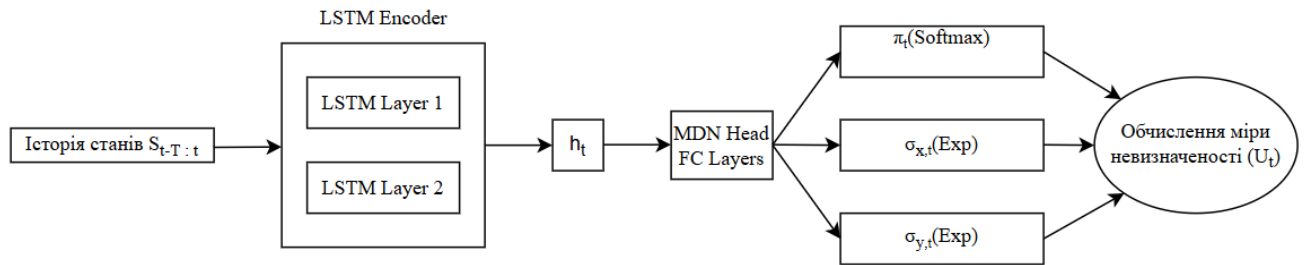


Рис. 3. 2. Структурна схема модуля ймовірнісного прогнозування

Запропонована модель дозволяє:

1. Генерувати мультимодальні прогнози траєкторій, враховуючи стохастичність людського руху.
2. Отримувати параметричний опис розподілу ймовірностей.

3. Розраховувати скалярну метрику невизначеності U_t аналітичним шляхом.

Метрика U_t є вхідним параметром для механізму адаптивного формування винагороди. Зазначений механізм забезпечує зв'язок між підсистемою прогнозування та підсистемою прийняття рішень.

3.4. Метод адаптивного формування функції винагороди на основі прогнозу невизначеності динамічного середовища

Ключовим елементом, який визначає ефективність навчання агента з підкріпленням, є функція винагороди. Для розробки ефективних RL-систем необхідно, щоб функція винагороди задовольняла таким властивостям, як інваріантність, інтерпретованість та інформативність [123].

Властивість інваріантності передбачає, що будь-яке перетворення або формування винагороди не повинно змінювати множину оптимальних політик. Дана властивість зберігає відповідність між розробленою винагородою та цілями задачі. За відсутності інваріантності робот може експлуатувати структуру винагороди і мати непередбачувану поведінку. Така поведінка робота відома як «злам винагороди».

Інтерпретованість визначає, наскільки легко розробник може зрозуміти функцію винагороди. Зазначена властивість забезпечує прозорість структури винагороди та її відповідність інтуїтивному людському розумінню поставленої задачі. Зокрема, у контексті соціально-орієнтованої навігації ця характеристика дає змогу чітко диференціювати внесок кожного компонента: від забезпечення фізичної безпеки до дотримання соціальних норм взаємодії.

У складних задачах робототехніки, в яких винагороди можуть визначатися через логіку або підцілі, інтерпретованість полегшує їх налагодження та верифікацію. Проектування інтерпретованих функцій винагороди стикається з компромісом між простотою та необхідністю деталізованого зворотного зв'язку. Розріджені функції винагороди покращують інтерпретованість, але мінімізують рівень деталізації, який необхідний для ефективного навчання. Одним із рішень

такої проблеми є використання структурних сигналів винагороди, які розбивають складні завдання на простіші підцілі, зберігаючи при цьому достатню деталізацію. Така декомпозиція дозволяє розробнику ідентифікувати конкретні компоненти моделі, які потребують корекції у разі виникнення небажаної поведінки агента.

Інформативність функції винагороди вимірює, наскільки ефективно вона надає корисні сигнали агенту для навчання та формування бажаної поведінки. Інформативна винагорода забезпечує послідовний зворотний зв'язок, який допомагає агенту швидко асоціювати дії з їхніми наслідками. Властивість важлива в середовищах із розрідженими винагородами, в яких зв'язок між діями та наслідками не є очевидним. Підвищення рівня інформативності безпосередньо корелює зі швидкістю збіжності алгоритму, оскільки насичений інформаційний потік мінімізує дисперсію градієнтів політики під час оновлення нейронної мережі.

Практична імплементація функції винагороди, яка б одночасно задовольняла вищезазначеним критеріям, є нетривіальним завданням, особливо в умовах стохастичного соціального середовища. Основна складність полягає не лише у формулюванні окремих компонентів винагороди, а й у виборі механізму їх інтеграції, який визначає пріоритетність цілей у кожний момент часу. Саме спосіб поєднання цих сигналів має вирішальний вплив на здатність агента знаходити компроміс між суперечливими вимогами ефективності та безпеки.

Аналіз існуючих підходів показав, що використання статичних вагових коефіцієнтів для компонентів функції винагороди призводить до суттєвих обмежень. З'являються дві діаметрально протилежні проблеми:

1. При високому значенні вагового коефіцієнта цільової ефективності агент може ігнорувати соціальний комфорт людини і рухатись надто близько або швидко біля неї;
2. При високих вагових коефіцієнтах безпеки серед людей агент може потрапити в локальний мінімум і зупинитись через неможливість знайти абсолютно безпечний шлях.

Для вирішення цих проблем запропоновано метод адаптивного формування функції винагороди [124].

Блок-схему методу адаптивного формування функції винагороди на основі прогнозування невизначеності сцени представлено на рис. 3.3.

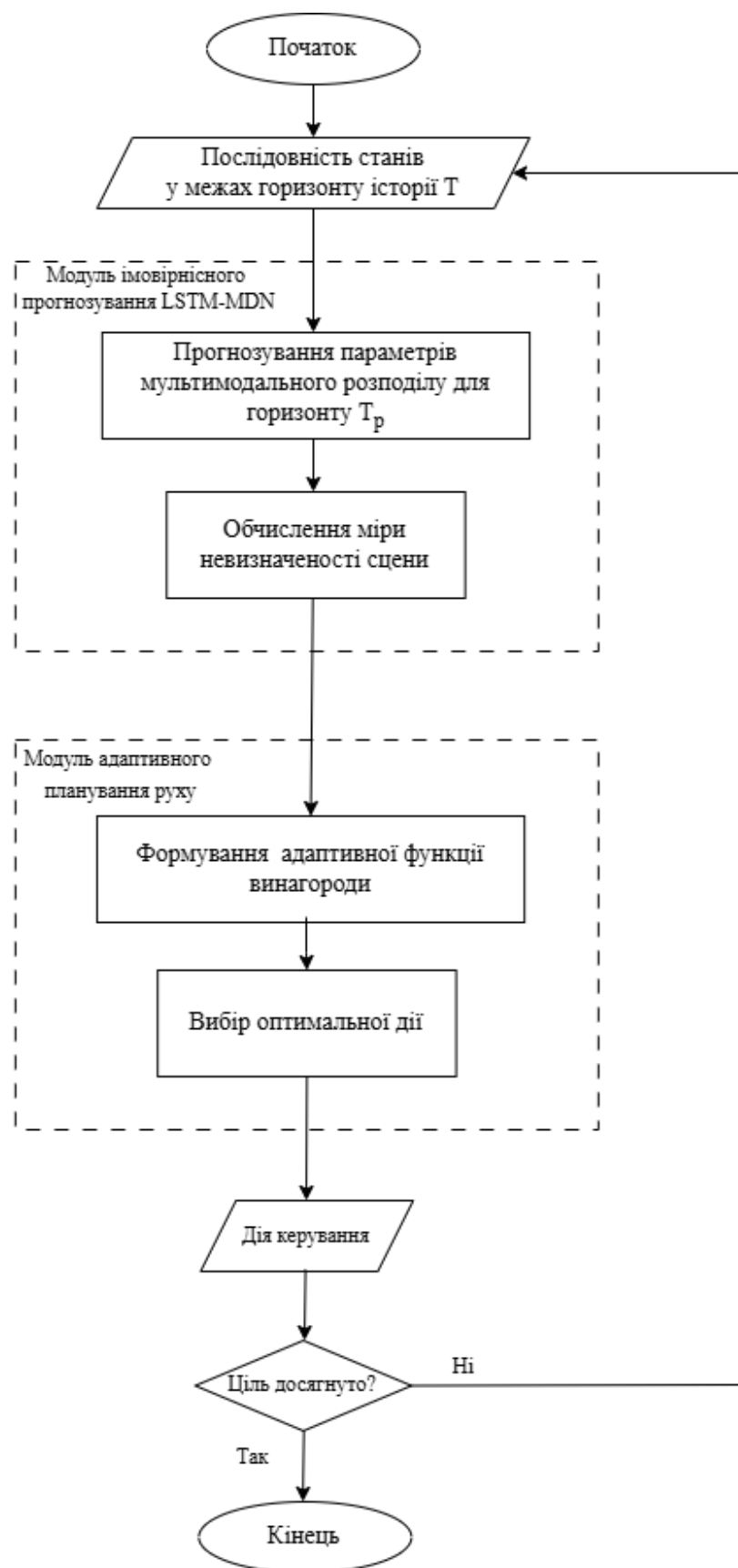


Рис. 3.3. Блок-схема методу адаптивного формування функції винагороди

Для реалізації системи соціально-адаптивної навігації створено гібридну архітектуру, яка складається з двох функціональних модулів:

1. Модуль імовірнісного прогнозування на основі LSTM-MDN забезпечує оцінку невизначеності середовища.
2. Модуль адаптивного планування руху на основі PPO, який використовує отриману оцінку невизначеності для адаптації функції винагороди та вибору оптимальної дії.

Така декомпозиція дозволяє мінімізувати затримки в контурі керування та забезпечити AMP диференційованим сигналом про ризики.

Структурну схему взаємодії AMP із середовищем у процесі навчання представлено на рис. 3. 4.

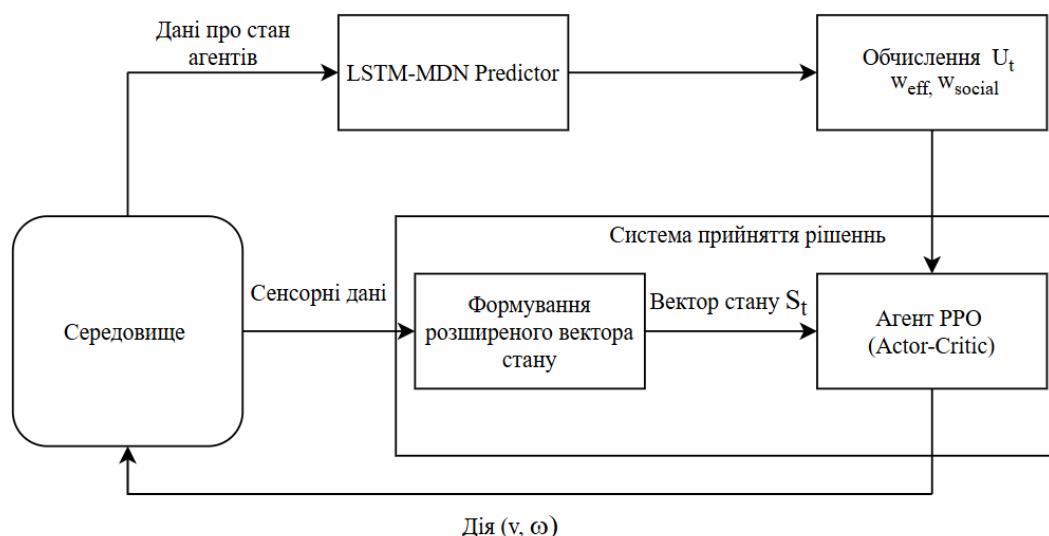


Рис. 3.4. Структурна схема взаємодії AMP із середовищем у процесі навчання

Запропонований метод динамічно коригує вагові коефіцієнти в залежності від рівня невизначеності навколишнього середовища.

Загальна функція винагороди R_t на кроці часу t , відповідно до описаної в розділі 2.4. моделі, визначається як сума компонентів:

$$R_t = R_{term,t} + R_{efficiency,t} + R_{social,t}. \quad (3.18)$$

Принцип функціонування розробленого методу полягає у впровадженні механізму динамічного зважування компонентів R_t на основі міри невизначеності U_t , отриманої від модуля імовірнісного прогнозування. У межах компонента

ефективності запропонованої моделі динамічному зважуванню підлягає складова винагороди за наближення до цілі, тоді як штраф за витрачений час фіксується як константа. Наслідком такого архітектурного рішення є збереження стійкого стимулу AMP до мінімізації часу під час навігації.

Модифікована функція винагороди набуває вигляду:

$$R_t = R_{term,t} + (w_{eff}(U_t)R_{progress,t} + R_{time,t}) + w_{social}(U_t)R_{social,t}, \quad (3.19)$$

де $w_{eff}(U_t)$, $w_{social}(U_t)$ – функції адаптованих вагових коефіцієнтів.

Коригування пріоритетів компонентів здійснюється залежно від поточного рівня невизначеності U_t із використанням сигмоїдальної функції:

$$w_i(U_t) = w_{base,i} + \left(\frac{2c_{scale}}{1 + e^{-k(U_t - U_{mid})}} - c_{scale} \right), \quad (3.20)$$

де $w_{base,i}$ – базове значення вагового коефіцієнта компонента; c_{scale} – коефіцієнт масштабування, який визначає діапазон зміни вагових коефіцієнтів; k – параметр крутизни сигмоїди для регулювання чутливості до змін; U_{mid} – порогове значення невизначеності.

Вибір сигмоїдальної функції для відображення міри невизначеності U_t у простір вагових коефіцієнтів винагороди зумовлений необхідністю забезпечення стабільності навчання агента PPO та нелінійності реакції системи на зміни середовища.

Доцільність використання саме сигмоїдальної функції обґрунтовується чотирма ключовими факторами:

1. Гладкість та диференційованість.
2. Фільтрація сенсорного шуму.
3. Обмеженість області значень.
4. Керованість крутизною переходу.

Гладкість та диференційованість є важливими характеристиками, оскільки алгоритм PPO, як і більшість методів класу актора-критика, базується на градієнтній оптимізації. Функція винагороди є частиною цільової функції, градієнт якої обчислюється для оновлення вагових коефіцієнтів нейронної мережі.

Застосування функцій із жорстким пороговим перемиканням унеможливило б дотримання умови гладкості цільової функції та коректного обчислення градієнта в критичних точках, спричинило б виникнення розривів першого. Процес навчання в цьому випадку було б дестабілізовано.

Сигмоїдальна функція є гладкою функцією і гарантує існування похідної на всій області визначення. Тому забезпечується стабільне поширення градієнта навіть у моменти перемикання пріоритетів робота.

Ефективна фільтрація сенсорного шуму досягається завдяки тому, що у зонах низької невизначеності $U_t \ll U_{mid}$ та високої невизначеності $U_t \gg U_{mid}$ похідна сигмоїди наближається до нуля.

Лінійна функція трансформації $w(U_t) = a U_t + b$ була б надто чутливою до незначних флуктуацій U_t , спричинених сенсорним шумом або похибками апроксимації MDN. Сигмоїдальна форма дозволяє ігнорувати малі збурення (шум) у зонах впевненості, змінюючи вагові коефіцієнти лише тоді, коли невизначеність наближається до критичного порогу U_{mid} .

Обмеженість області значень відіграє вирішальну роль для стабільного навчання RL-агента, оскільки критично важливо, щоб компоненти винагороди залишалися у фіксованому діапазоні. Необмежене зростання вагових коефіцієнтів, яке можливе при лінійній або експоненційній залежності, може призвести до проблеми «вибуху градієнтів». Тому запропонована формула гарантує, що значення вагових коефіцієнтів завжди буде в чітко визначеному діапазоні

$$w_i(U_t) \in [w_{base,i} - c_{scale}, w_{base,i} + c_{scale}]. \quad (3.21)$$

Також обмеженість області значень дозволяє аналітично задавати межі поведінки робота, наприклад, максимальний штраф за небезпеку, запобігаючи надмірно консервативній поведінці.

Керованість крутизною переходу реалізується за допомогою параметра k , який дозволяє налаштовувати інтенсивність адаптації. При високих k система працює як м'який перемикач, різко змінюючи пріоритети при перетині порогу U_{mid} . При низьких k перехід стає більш плавним. Це дозволяє налаштувати реакцію

автономного робота під конкретну динаміку середовища без зміни архітектури нейронної мережі.

Таким чином, використання сигмоїдальної функції є оптимальним вибором, який гарантує стабільність оптимізації політики методом градієнтного спуску. Водночас цей підхід забезпечує ефективну нелінійну адаптацію навігаційної поведінки автономного робота за умов стохастичної невизначеності середовища.

Розроблений метод адаптивного формування функції винагороди дозволяє досягти оптимального співвідношення між безпекою, ефективністю та соціальною прийнятністю під час навігації автономних мобільних роботів. Використання оцінки невизначеності дозволяє АМР динамічно змінювати стиль поведінки в режимі реального часу та підвищувати загальну надійність системи в неструктурованих динамічних середовищах.

Висновки до розділу 3

У третьому розділі здійснено розробку та обґрунтування методу адаптивного формування винагороди для автономних мобільних роботів на основі оцінки стохастичної невизначеності динамічного середовища. Отримані результати дозволяють сформулювати такі висновки:

Здійснено обґрунтування архітектурних рішень для побудови інтелектуальної системи керування в соціальному середовищі. Для забезпечення роботи в режимі реального часу доцільно використовувати гібридну архітектуру, яка розділяє контури імовірнісного прогнозування траєкторій та адаптивного планування руху.

Розроблено математичну модель імовірнісного прогнозування станів динамічних об'єктів на основі поєднання рекурентних нейронних мереж (LSTM) та мереж суміші густин (MDN). Використання рекурентних нейронних мереж дає змогу ефективно враховувати часові залежності, тоді як компонент мережі суміші густин генерує набір гаусових розподілів. Запропонована архітектура забезпечує апроксимацію мультимодального розподілу вірогідних просторових положень

агентів-людей для подальшого аналітичного обчислення міри невизначеності середовища.

Запропоновано метод адаптивного формування функції винагороди, який на основі змін середовища забезпечує зміну навігаційної стратегії робота. Розроблений метод передбачає динамічну адаптацію вагових коефіцієнтів функції винагороди на основі обчисленої міри невизначеності середовища, забезпечуючи оптимальний баланс між навігаційною ефективністю, безпекою та соціальною прийнятністю. Зокрема, в умовах високої невизначеності динамічного середовища система автоматично збільшує вагу штрафів за порушення соціальної дистанції, змушуючи АМР обирати консервативну поведінку, а в середовищі з відсутністю людей максимізує швидкість досягнення цілі.

Таким чином, у розділі розроблено математичний та алгоритмічний апарат, який дозволяє реалізувати механізм адаптивного формування винагороди для навчання автономного мобільного робота відповідно до змін стохастичного динамічного середовища. Створене теоретичне підґрунтя є основою для подальшої програмної реалізації та верифікації навігаційної системи у симуляційних середовищах із застосуванням алгоритмів глибинного навчання з підкріпленням.

РОЗДІЛ 4

ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

4.1. Розробка симуляційного середовища для валідації методів соціально-адаптивної навігації

Експериментальна перевірка та оцінка ефективності запропонованих методів соціально-адаптивної навігації зумовлюють необхідність створення середовища для апробації та тестування розроблених рішень. Безпосереднє розгортання неапробованих алгоритмів керування в динамічних соціальних середовищах стає причиною високої ймовірності виникнення критичних зіткнень. Методи навігації АМР потребують попередньої верифікації в симуляційних середовищах.

Для ітеративного відпрацювання механізмів взаємодії мобільного робота із перешкодами та конфігурування політик було розроблено 2D-симуляційне середовище.

Структурну схему модулів симуляційного середовища представлено на рис. 4. 1.

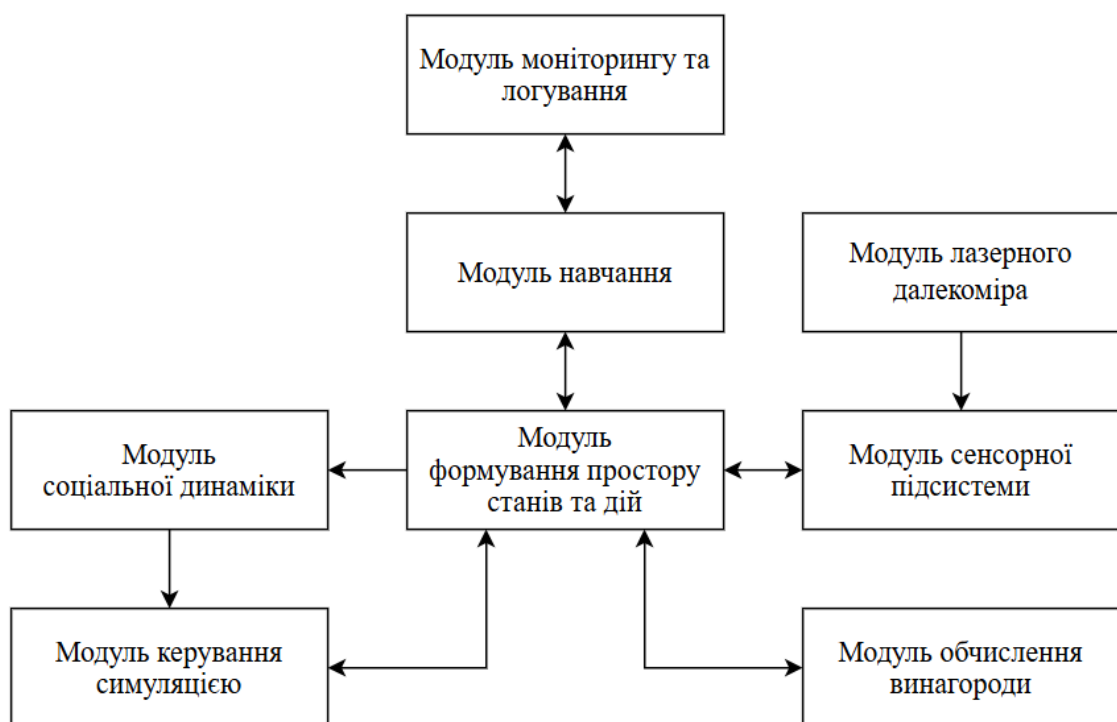


Рис. 4.1. Структурна схема модулів симуляційного середовища.

Архітектура програмного рішення побудована на базі стандартизованого інтерфейсу `gymnasium.Env`, який забезпечує інтеграцію з сучасними бібліотеками глибокого навчання з підкріпленням. Програмний комплекс реалізований засобами мови програмування Python та передбачає методологічне спрощення динамічних характеристик твердих тіл та сил тертя, зосереджуючись на кінематиці переміщення й аналізі просторових метрик. Такий підхід дозволяє збільшити частоту генерації навчальних епізодів під час тензорних обчислень, забезпечуючи умови для оперативного налаштування гіперпараметрів та структури функції винагороди в процесі глибокого навчання з підкріпленням. Архітектура програмного комплексу розподілена на модулі.

Модуль формування простору станів та дій `social_env` є інтеграційним ядром системи. Він здійснює фінальну конфігурацію простору неперервних дій перед передачею даних алгоритмам глибокого навчання.

Простір дій робота представлений двовимірним вектором команд керування. Встановлені діапазони значень у межах простору дій визначено фізичними обмеженнями цільової апаратної платформи та визначають лінійну швидкість у діапазоні від 0,0 до 0,26 м/с та кутову швидкість – у діапазоні від -1,5 до 1,5 рад/с.

Структура простору станів охоплює такі компоненти:

- відносна відстань та відносний кут напрямку до цільової точки у локальній системі координат робота;
- поточні значення лінійної та кутової швидкостей мобільної платформи;
- масив даних лазерного далекоміра, який складається з 360 променів із максимальною дальністю вимірювання 3,5 метра та математично моделює процес трасування променів до найближчих перешкод.

Модуль лазерного далекоміра `lidar_modul` виконує функцію імітаційного моделювання роботи лідара. Зазначений програмний компонент реалізує векторизований алгоритм розрахунку точок перетину віртуальних променів із наявними перешкодами.

Застосування матричних операцій дозволяє системі одночасно обчислювати відстані для всього масиву із 360 променів. Такий підхід знижує обчислювальне

навантаження під час генерації навчальних епізодів. Знайдені відстані до найближчих об'єктів формують скан простору, який передається до модуля сенсорної підсистеми для подальшого накладання стохастичних шумів.

Модуль сенсорної підсистеми `sensor_module` відповідає за релевантне відтворення апаратних особливостей сприйняття та формування багатовимірного масиву спостережень.

Ефективність перенесення політики керування, сформованої в ітераційному процесі навчання з підкріпленням, безпосередньо залежить від ступеня відповідності віртуального середовища фізичним реаліям. 2D-симулятори часто характеризуються надмірною детермінованістю та ідеалізацією фізичних процесів, наслідком яких є перенавчання агента на математичних артефактах моделі. Для подолання зазначеного розриву у розроблене середовище навчання впроваджено механізм моделювання стохастичних процесів та факторів невизначеності.

Важливим етапом побудови робастного середовища симуляції є формалізація стохастичних похибок сенсорної та кінематичної підсистем. Врахування апаратних похибок реальних сенсорів реалізовано шляхом внесення стохастичних збурень у вектор спостережень. Параметри шумів у розробленому двовимірному середовищі було встановлено відповідно до офіційних технічних характеристик робота TurtleBot3 Waffle та його конфігураційних файлів Simulation Description Format (SDF) симулятора Gazebo [125, 126].

Для моделювання роботи лазерного далекоміра (LiDAR) застосовано адитивний білий гаусовий шум (AWGN), який імітує випадкові коливання вимірювань, спричинені тепловими шумами фотоприймача та дискретністю обробки сигналу. Математична модель вимірювання відбиття променя z_t на кожному кроці симуляції описується виразом

$$z_{t,i} = \max\left(0, \min(z_{t,i}^* + N(0, \sigma_{lidar}^2), z_{max})\right), \quad (4.1)$$

де $z_{t,i}^*$ – відстань до перешкоди для i -го променя, отримана шляхом геометричного трасування; σ_{lidar} – середньоквадратичне відхилення, встановлене на рівні 0,015 м

відповідно до метрологічних характеристик LiDAR-систем середнього класу; Z_{max} – максимальний діапазон сенсора.

Окрему увагу приділено випадковій втраті вимірювань LIDAR. Враховуючи фізичні властивості поверхонь (дзеркальність, високе поглинання), впроваджено функцію варіативного заміщення значень променів максимальним діапазоном Z_{max} із параметром імовірності $p = 0,01$. Це стимулює нейромережу формувати стратегії уникнення перешкод, стійкі до короткочасної відсутності сигналу. Похибки одометрії, які виникають через проковзування коліс та неточності енкoderів, моделюються як збурення актуальних лінійної та кутової швидкостей із відхиленнями $\sigma_v = 0,01$ м/с та $\sigma_\omega = 0,02$ рад/с. Програмна реалізація механізму стохастичного шуму сенсорних даних представлена у лістингу 4. 1.

Лістинг 4. 1. Додавання стохастичного шуму

```
def _apply_sensor_noise(self, ideal_laser, ideal_vel):
    # Накладання гаусового шуму на дані LiDAR
    noisy_laser = ideal_laser + np.random.normal(0, 0.015,
size=self.num_laser_beams)

    # Моделювання втрати вимірювань
    dropout_mask = np.random.rand(self.num_laser_beams) < 0.01
    noisy_laser[dropout_mask] = self.max_laser_range
    noisy_laser = np.clip(noisy_laser, 0.0, self.max_laser_range)

    # Збурення швидкості для імітації похибок одометрії
    noisy_vel = ideal_vel + np.random.normal(0, [0.01, 0.02])

    return noisy_laser, noisy_vel
```

Наступний аспект підвищення реалістичності симуляції охоплює моделювання латентності даних та часових затримок. Однією з проблем при інтеграції DRL-агентів у реальні робототехнічні системи є часова затримка між моментом фізичного зчитування даних та надходженням вектора станів на вхід нейронної мережі. У сучасних архітектурах на базі ROS 2 [127] латентність обумовлена асинхронністю публікації топиків, ресурсними витратами проміжного програмного забезпечення (DDS) та часом обробки переривань центральним процесором [128,129].

Для імітації зазначеного ефекту у програмному середовищі реалізовано механізм часової затримки через структуру даних «черга» (FIFO buffer queue) з

фіксованою глибиною L . За умови кроку дискретизації Δt , еквівалентна латентність даних LiDAR становить

$$T_{delay} = L\Delta t. \quad (4.2)$$

Такий підхід змінює динаміку навчання і алгоритм адаптується до прийняття рішень на основі відтермінованих за часом даних. Дана методологія детермінує здатність агента до імпліцитного вивчення динаміки руху об'єктів та формування стратегій випереджаючого керування.

Функціональне призначення модуля соціальної динаміки `social_dynamics_module` полягає у відтворенні реалістичного руху агентів-людей. Реалізація соціальних взаємодій у 2D-просторі базується на механізмі просторового зонування та випадковій генерації автономних агентів-людей. Для моделювання динаміки агентів-людей використано модель соціальних сил (SFM), згідно з якою рух кожного об'єкта визначається векторною сумою сил притягання та відштовхування [130]. Алгоритм ітеративно обчислює взаємодію із статичними перешкодами та мобільним роботом. Ітеративна процедура детермінує виникнення складних поведінкових патернів і дозволяє створити середовище з високим рівнем стохастичної невизначеності.

З метою формалізації соціально прийнятної поведінки робота у програмному коді визначено проксемічні зони навколо кожної людини-агента.

Модуль керування симуляцією `simulation_module` виконує роль контролера імітаційного процесу. В модулі впроваджено механізм дворівневої дискретизації часу. Процес прийняття рішень інтелектуальним агентом синхронізовано з макрокроком тривалістю 0,25 с, задаючи частоту звернень до нейромережевої політики. Для підвищення реалістичності взаємодії впроваджено внутрішній фізичний цикл (мікрокрок) тривалістю 0,05 с. Протягом кожного мікрокроку виконується ітеративний перерахунок позицій об'єктів на основі сигналів керування, забезпечуючи плавну трансформацію просторових координат усіх об'єктів сцени та релевантне моделювання динамічних ефектів.

Модуль здійснює геометричний аналіз напрямку руху. Цей компонент обчислює косинус кута між вектором поточного курсу робота та вектором напрямку руху людини. Такий математичний апарат дозволяє системі ідентифікувати соціально неприйнятні ситуації фронтального зближення.

Модуль обчислення винагороди `reward_module` перетворює просторову конфігурацію об'єктів у скалярний сигнал зворотного зв'язку. Система агрегує показники прогресу руху до цілі та штрафи за недотримання соціальних норм.

Модуль моніторингу та логування `train_callbacks` призначений для супроводу процесу навчання та фіксації даних дослідження. Здійснюється експорт метрик у формати CSV та TensorBoard для подальшого аналізу збіжності алгоритму.

Модуль навчання `train` ініціалізує обчислювальні ресурси та параметри DRL-алгоритму. Даний модуль координує роботу всіх підсистем, запускає ітераційний цикл взаємодії агента із середовищем.

Для здійснення візуального контролю за процесом навчання AMP розроблено підсистему рендерингу. Вона базується на інструментарії бібліотеки комп'ютерного зору OpenCV. Інтерфейс симулятора динамічно відображає стан середовища у двовимірній площині (рис. 4. 2).

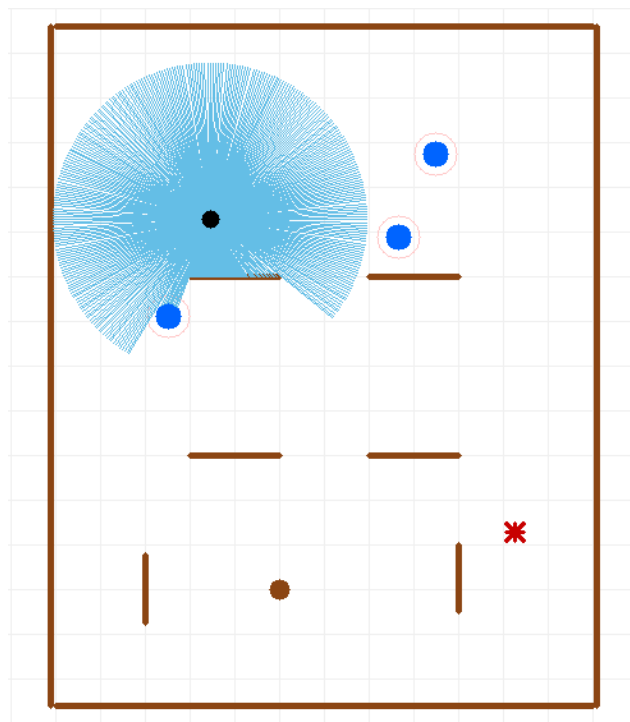


Рис. 4.2. Візуалізація сцени

Навігаційний простір, розміром 12×15 м, обмежено контурами стін. У межах простору розміщено цільову точку, яка позначається графічним маркером червоного кольору. Статичні перешкоди позначено коричневим кольором.

Мобільна платформа зображено чорним кругом, радіус якого $r = 0,22$ м. Від центру мобільної платформи розходяться промені лазерного далекоміра. Сукупність цих променів ілюструє поточну зону видимості сенсорної підсистеми робота та демонструє місця перетину променів із перешкодами.

Динамічні об'єкти представлені подвійною фігурою. Внутрішня фігура, синій круг, відповідає фізичним габаритам людини. Навколо неї – червоне коло, яке візуалізує межі особистого простору.

Таким чином, створене симуляційне середовище імітаційного моделювання становить цілісну систему для безпечної та ітеративної апробації алгоритмів соціальної навігації. Поєднання модулів дозволяє отримати робастну модель керування АМР, яка підвищує стабільність при перенесенні у фізичні симулятори або на реальну апаратну платформу.

4.2. Система метрик оцінювання ефективності соціально-адаптивної навігації

Об'єктивне порівняння алгоритмів навігації в умовах динамічного соціального середовища вимагає застосування комплексного підходу до оцінки згенерованих траєкторій. З огляду на специфіку взаємодії автономного мобільного робота з людьми, традиційних критеріїв недостатньо. Тому в дослідженні розроблено та застосовано багатокритеріальну систему оцінювання, яка складається з таких груп метрик:

1. Показники загальної успішності та надійності навігації.
2. Метрики ефективності планування просторових траєкторій.
3. Оцінка соціальної прийнятності маневрів.

Група критеріїв показників загальної успішності та надійності навігації характеризує базову здатність алгоритму безпечно виконувати поставлене завдання та включає наступні параметри:

- коефіцієнт успішності (Success Rate) – відсоткове відношення епізодів, під час яких АМР успішно досяг цільової точки без жодних зіткнень із перешкодами та в межах відведеного часу, до загальної кількості тестових запусків;
- кількість зіткнень із людьми (Human Collisions) – частота зіткнення з динамічним агентом;
- кількість зіткнень зі статичними об'єктами (Static Collisions) – частота зіткнень із стаціонарними перешкодами;
- кількість тайм-аутів (Timeouts) – показник частоти епізодів, в яких АМР не сформував безпечну траєкторію до цільової позиції в межах заданого часового інтервалу, слугує безпосереднім індикатором прояву проблеми «застиглого робота».

Категорія метрик ефективності планування просторових траєкторій дозволяє оцінити швидкісні та геометричні характеристики згенерованих маршрутів і включає такі метрики [131]:

- час руху (Time) – середній час у секундах, витрачений АМР на переміщення від стартової позиції до заданої цільової позиції;
- довжина траєкторії руху (Path) – середня фактична відстань у метрах, яку подолав АМР. В контексті соціальної навігації цей показник аналізується у зв'язці з часом виконання, оскільки безпечний превентивний обхід динамічних агентів часто вимагає побудови довшої, але більш ефективної в часі траєкторії.

Метрики оцінки соціальної прийнятності маневрів є ключовим для соціальної робототехніки і кількісно вимірює рівень комфорту, який робот створює для оточуючих людей, за допомогою наступних метрик:

- узагальнений показник відповідності соціальним нормам (Social Compliance Score, SCS) – зведений показник у діапазоні від 0% до 100%,

який відображає загальний рівень дотримання роботом соціальних норм під час навігації;

- час перебування в інтимному просторі – тривалість перебування робота на критично небезпечній відстані від людини (менше 0.5 м);
- час перебування в особистому просторі – тривалість перебування мобільної платформи в радіусі до 1 м від агента-людини, що викликає помірний психологічний дискомфорт;
- час фронтального зближення – тривалість руху робота при фронтальному зближенні.

Для кількісної оцінки рівня дотримання АМР соціальних норм в роботі запропоновано узагальнений показник відповідності соціальним нормам (SCS).

Необхідність введення даного показника зумовлена специфікою задачі соціальної навігації, успішність якої визначається не лише фактом досягнення цільової точки, а й мінімізацією негативного впливу на людей. На відміну від дискретних метрик (наприклад, кількості зіткнень), SCS базується на часі перебування в проксемічних зонах. Це дозволяє врахувати «накопичувальний» ефект дискомфорту, який відчуває людина при тривалому перебуванні робота в її особистому просторі.

Показник SCS розраховується у відсотковому діапазоні від 0 до 100 % та базується на зваженій сумі часу порушень проксемічних зон відносно загального часу руху. Оцінка SCS визначається за формулою

$$SCS = \max \left(0, \left(1 - \frac{W_I \cdot T_{int} + W_P \cdot T_{per} + W_F \cdot T_{frt}}{T_{ref}} \right) \cdot 100 \right), \quad (4.4)$$

де T_{int} – час перебування робота в інтимному просторі людини (відстань менше 0.5 м); T_{per} – час перебування робота в особистому просторі людини (відстань від 0.5 м до 1.0 м); T_{frt} – час руху робота назустріч людині (фронтальне зближення); T_{ref} – еталонний час руху робота від початкової точки до цілі за умови руху по прямій із максимальною швидкістю V_{max} ; W_I, W_P, W_F – вагові коефіцієнти.

Вагові коефіцієнти W_I , W_P , W_F дозволяють адаптувати метрику під різні сценарії тестування. Наприклад, суворіші вимоги для медичних закладів або лояльніші для складських приміщень з навченим персоналом.

Максимальне значення 100% відповідає високому рівню соціальної адаптивності. Даний результат свідчить про формування роботом такої траєкторії руху, при якій протягом усього навігаційного епізоду не було зафіксовано жодного факту перетину інтимного чи особистого простору проксеміки людини, а також не виявлено випадків потенційно конфліктного фронтального зближення.

Мінімальне значення 0% визначає критично низький рівень соціальної адаптивності або повну невідповідність поведінки АМР заданим обмеженням. Такий показник фіксується у випадках, коли сумарний зважений час порушень зон проксеміки та фронтальних наближень дорівнює часу руху.

Застосування функції максимуму гарантує, що у випадку критичної кількості порушень метрика не набуватиме від'ємних значень, забезпечуючи коректність статистичного аналізу під час порівняння різних навігаційних алгоритмів.

Таким чином, сформовано багатокритеріальну систему оцінювання, яка поєднує технічні метрики успішності навігації з показниками соціальної прийнятності АМР.

4.3. Експериментальне дослідження методу навчання навігаційної політики на основі глибинного навчання з підкріпленням та Curriculum Learning

Експериментальне дослідження спрямоване на валідацію запропонованого методу соціально-адаптивної навігації автономних мобільних роботів у динамічному середовищі. Основна увага приділяється оцінці результативності інтеграції алгоритму PPO зі стратегією Curriculum Learning (PPO-CL).

Реалізація серії експериментів у спеціально розробленому 2D-середовищі має на меті перевірити гіпотезу: поступове ускладнення параметрів середовища разом із поетапним розширенням компонентів функції винагороди формують

стійку політику керування. Основне завдання полягає в оптимізації навігації АМР шляхом встановлення оптимального співвідношення між безпечним й ефективним рухом та дотриманням норм соціально прийнятної поведінки.

Для валідації запропонованого методу здійснюється порівняльний аналіз ефективності навчання агента за допомогою комплексного підходу PPO-CL порівняно з алгоритмом PPO без застосування CL.

Навчання інтелектуального агента тривало протягом 1 мільйона кроків взаємодії (timesteps). Для навчання нейронної мережі було застосовано алгоритм PPO, реалізованого на базі бібліотеки Stable Baselines3 із застосуванням архітектури багатошарового перцептрона для політики та функції цінності.

Гіперпараметри алгоритму PPO:

Коефіцієнт навчання (Learning Rate) – 0,0002.

Кількість кроків у буфері (n_steps) – 8192.

Розмір міні-батчу (Batch Size) – 2048.

Коефіцієнт дисконтування (γ) – 0.99.

Параметр згладжування GAE (λ) – 0.95.

Коефіцієнт ентропії – 0.015.

Процес навчання методом на основі PPO та Curriculum Learning здійснювався за сценарієм поступового ускладнення середовища та модифікації функції винагороди. Перехід між етапами відбувався автоматично за умови досягнення агентом встановленого рівня успішності та виконання критеріїв стаціонарності та стабільності навчання. Програма навчання охоплювала чотири послідовні етапи навчання, які детально описано в розділі 2. 6.

У таблиці 4. 1 вказано розмір ковзного часового вікна N_l та порогові значення $\varepsilon_l, \delta_l, \eta_l$ для відповідних етапів навчання.

Таблиця 4.1.

Параметри переходу між етапами навчання

Етап	N	ε	δ	η
Етап 2	40 000	0,2	4	75
Етап 3	80 000	0,4	6	75
Етап 4	100 000	0,6	7	75

Зафіксовані під час експерименту фактичні показники збіжності на момент переходу представлено у таблиці 4.2.

Таблиця 4.2.

Фактичні показники збіжності за етапами Curriculum Learning

Етап переходу	Глобальний крок	Фактична зміна середньої винагороди, $\Delta \bar{J}$	Стандартне відхилення, σ	Рівень успішності, S
Етап 2	188416	0,1461	3,8945	89,14
Етап 3	348672	0,2247	4.1092	90,01
Етап 4	548864	0,1120	4,4109	88,35

У межах дослідження було встановлено, що безпосереднє впровадження соціального компонента функції винагороди при переході на 3 етап навчання ускладнює процес оптимізації. Для подолання цієї проблеми у роботі впроваджено механізм відпалу соціального компонента винагороди.

Механізм відпалу реалізовано через параметр α_t , який виступає динамічним множником для соціальної складової винагороди агента і реалізується через множник $\alpha_t \in [0,1]$

$$R_t = R_{\text{term},t} + R_{\text{efficiency},t} + \alpha_t R_{\text{social},t} , \quad (4.5)$$

де коефіцієнт відпалу α_t у програмній реалізації представлено як кусково-лінійну функцію. Її значення залежить від поточного етапу навчальної програми Curriculum Learning та кількості виконаних кроків навчання

$$\alpha_t = \begin{cases} 0, & \text{при } S \in \{1, 2\}, \\ \min\left(1; \frac{t - t_{\text{start}}}{N_{\text{anneal}}}\right), & \text{при } S = 3, \\ 1, & \text{при } S \geq 4, \end{cases} \quad (4.6)$$

де S – ідентифікатор поточного етапу CL; t – поточний глобальний крок симуляції; t_{start} – крок початку третього етапу навчання; N_{anneal} – фіксована тривалість фази відпалу.

Під час експериментального дослідження було використано значення $N_{\text{anneal}} = 80000$.

Результати тестування методу PPO-CL з відпалом та без відпалу продемонстрували збільшення успішних епізодів із 93% до 95%, що свідчить про позитивний вплив механізму відпалу соціальної складової функції винагороди на процес навчання агента.

Зокрема, у варіанті PPO-CL без відпалу спостерігався нижчий рівень збіжності кумулятивної винагороди. Це пояснюється тим, що соціальні штрафи застосовуються одразу із повною інтенсивністю і ускладнюють дослідження простору станів. Натомість у запропонованому підході PPO-CL із відпалом соціальної винагороди навчання відбувалося більш стабільно та ефективно.

На рисунку 4. 3. представлено графіки кумулятивних винагород, отриманих у процесі навчання моделей із застосуванням відпалу та без відпалу.

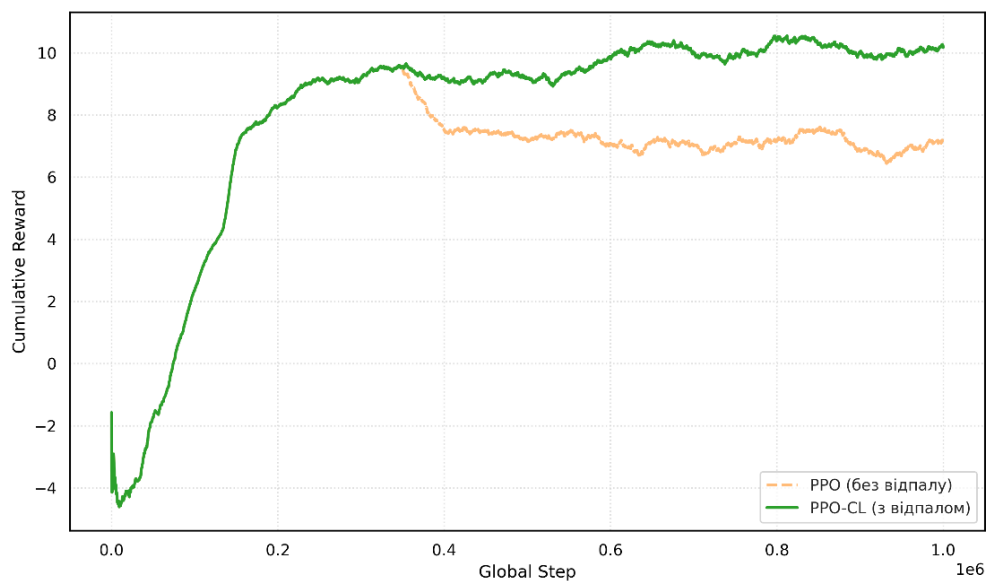


Рис. 4.3. Графік кумулятивних винагород навчання моделей із відпалом та без відпалу

Таким чином, застосування механізму відпалу дозволив зберегти високий рівень кумулятивної винагороди та підвищити стабільність політики.

На рисунку 4. 4 представлено результат процесу навчання моделей на основі методу PPO та запропонованого підходу PPO-CL з реалізацією відпалу. Аналіз проведено за відсотком успішних епізодів залежно від кількості кроків навчання.

Рисунок 4. 4 демонструє, що запропонований метод PPO-CL має суттєво швидшу динаміку навчання порівняно з алгоритмом PPO.

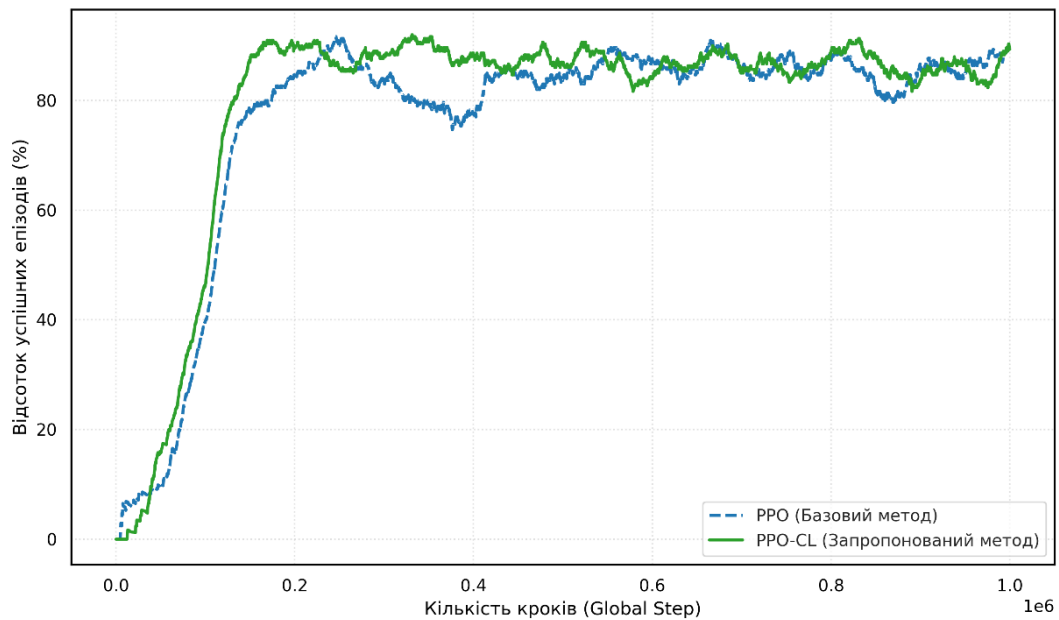


Рис. 4.4. Відсоток успішних епізодів залежно від кількості кроків навчання для моделей PPO та PPO-CL

На початкових етапах навчання спостерігається стрімке зростання частки успішних епізодів для PPO-CL, яка досягає рівня близько 90%. Натомість PPO характеризується повільнішим зростанням цього показника. Подальший аналіз кривих свідчить про поступову стабілізацію обох методів, однак PPO-CL досягає стаціонарного режиму значно раніше. У фінальній фазі навчання обидва підходи демонструють подібний рівень ефективності, який знаходиться в межах 85–90% успішних епізодів.

Таким чином, результати свідчать про те, що запропонований підхід PPO-CL забезпечує суттєве покращення швидкості навчання та стабільності процесу оптимізації політики порівняно з методом PPO.

Для оцінювання ефективності навчених політик керування застосовано комплекс кількісних метрик, зазначених в розділі 4. 2. Під час аналізу отриманих даних головна увага приділялася показникам успішності та безпеки навігації, ступеню дотримання роботом соціальних норм, а також загальним просторово-часовим і кінематичним характеристикам руху автономної платформи.

Результати порівняльного аналізу ефективності методу PPO-CL та методу PPO представлені на основі валідаційного тестування у чотирьох тестових зонах із

підвищенням рівня складності. Перша тестова зона моделює базове середовище, облаштоване виключно статичними перешкодами різної конфігурації. Другий етап ускладнює завдання шляхом інтеграції одного динамічного агента-людини у межах існуючого статичного середовища. Третя зона передбачає взаємодію автономної платформи з двома рухомими людьми та із статичними об'єктами. Четверта тестова зона репрезентує просторово масштабоване комплексне середовище та акумулює параметри всіх попередніх етапів: одночасний рух трьох агентів-людей, збільшення площі середовища тестування та кількості статичних перешкод.

Аналіз результатів валідаційного тестування засвідчує загальну перевагу комплексного підходу PPO-CL за критерієм успішності досягнення цілі. Інтегрований метод продемонстрував загальний показник успішності на рівні 95% порівняно з 87% для базового алгоритму PPO (рис. 4. 5).

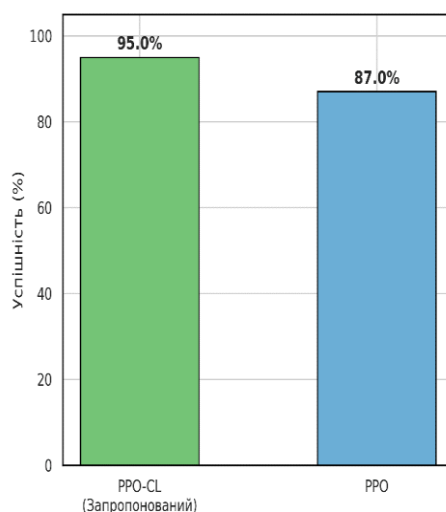


Рис. 4.5. Порівняльний аналіз показника успішно завершених епізодів.

Окремої уваги заслуговує аналіз кількості зіткнень із динамічними об'єктами та статичними перешкодами, результати якого відображено на рисунку 4.6. Метод PPO-CL має значно вищий рівень безпеки соціально-адаптивної навігації. Загальна кількість зіткнень з людьми зменшилася в 2 рази. Кількість зіткнень зі статичними перешкодами також знизилася з 9 до 3 на перевагу інтегрованого методу.

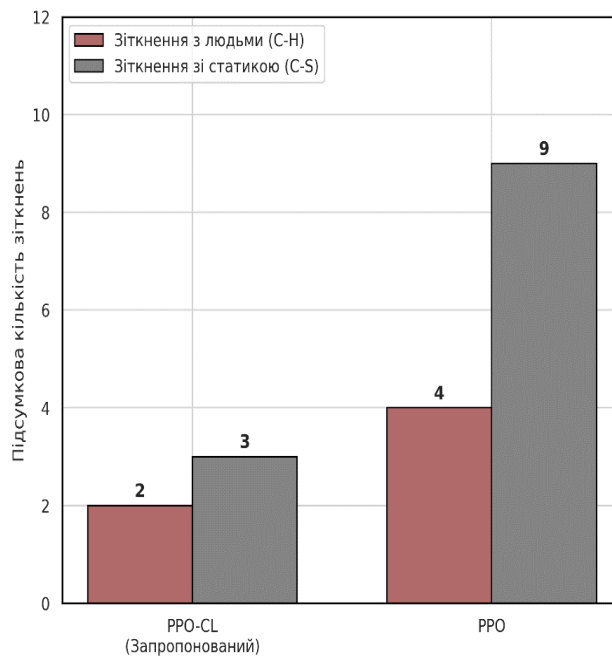


Рис. 4.6. Порівняльний аналіз кількості зіткнень із динамічними об'єктами та статичними перешкодами

Отримані результати емпірично підтверджують здатність стратегії Curriculum Learning формувати надійні механізми обходу перешкод із фокусом на безпеці людини.

Порівняльний аналіз узагальненого показника відповідності соціальним нормам (SCS) ілюструє переваги застосування стратегії Curriculum Learning для формування безпечної поведінки автономного мобільного робота.

Інтегрований підхід PPO-CL забезпечив підвищення SCS до 72,3% порівняно з 72,0% для алгоритму PPO. Ускладнення просторових умов шляхом збільшення кількості рухомих агентів (третя зона) закономірно призводить до зниження метрики для обох підходів через підвищену щільність динамічних перешкод. Водночас політика PPO-CL демонструє вищу стійкість, зберігаючи показник на рівні 66,7%, тоді як результативність PPO знижується до 65,7%. Наведені дані кількісно підтверджують здатність поетапного навчання формувати навички уникнення порушень проксемічних зон людини в умовах тісної взаємодії.

Найбільш репрезентативні результати зафіксовано у комплексному середовищі (Full Map), де показник SCS для запропонованого методу склав 76,9%, що значно перевищує результат базового алгоритму 70,7%. Зафіксоване зменшення

тривалості взаємодії на критично близьких відстанях безпосередньо корелює зі зміною загальних кінематичних показників навігації. Порівняльний аналіз виявляє закономірне зростання загального часу проходження маршруту (з 21,58 с до 22,65 с) та збільшення довжини пройденого шляху (з 5,34 м до 5,68 м) при впровадженні стратегії CL. Даний результат має чітке наукове обґрунтування в контексті завдань соціально-адаптивної навігації.

Забезпечення зростання показника метрики SCS та мінімізація кількості небезпечних зближень з людьми вимагають від АМР виконання превентивних маневрів ухилення. Побудова обхідних траєкторій, розрахованих на уникнення проксемічних зон людини по ширшому радіусу, математично та фізично зумовлює відхилення від найкоротшого геометричного маршруту. З огляду на це, виявлене збільшення навігаційного часу та пройденої відстані розглядається як об'єктивно необхідний компроміс між абсолютною кінематичною ефективністю та безпекою функціонування мобільного робота в динамічному середовищі.

Порівняльний аналіз ефективності функціонування розглянутих алгоритмічних рішень узагальнено в таблиці 4.3.

Таблиця 4.3

Порівняння показників ефективності

Метрика	PPO-CL	PPO	Динаміка змін
Success Rate, %	95	87	+8
Collisions, %	5	13	-8
Human Collisions, %	2	4	-1
Static Collisions, %	3	9	-5
Timeouts, %	0	0	0
Time, с	22,65	21,58	+0,59
Path, м	5,68	5,34	+0,2
SCS, %	72,3	72,0	+0,3

Проведене експериментальне дослідження та комплексний аналіз кількісних показників дозволяють зробити висновок стосовно ефективності запропонованого алгоритму PPO-CL для вирішення завдань соціально-адаптивної навігації.

Інтеграція стратегії Curriculum Learning у архітектуру PPO забезпечує вищі показники ефективності моделі. Розроблений гібридний метод демонструє підвищення загальної успішності виконання навігаційних епізодів з 87 % до 95 %.

Ключовою перевагою моделі PPO-CL є підвищення рівня безпеки та соціальної прийнятності згенерованих траєкторій. Загальна кількість зіткнень зменшується на 8 %, включно зі зниженням рівня критичних зіткнень з людьми. Зафіксоване в ході експериментів незначне зростання часових та просторових витрат на проходження маршруту є необхідним компромісом. Здатність навігаційної системи завчасно планувати безпечні обхідні маневри виправдовує відхилення від найкоротших шляхів. Отже, застосування методу PPO-CL дозволило створити надійну, соціально орієнтовану систему керування автономним мобільним роботом для функціонування в динамічному людському середовищі.

4.4. Оцінка ефективності методу адаптивного формування винагороди на основі прогнозу невизначеності

Метою даного дослідження є кількісний та якісний аналіз ефективності запропонованого методу адаптивного формування винагороди на основі прогнозу невизначеності. Запропонований підхід базується на інтеграції алгоритму PPO, рекурентної нейронної мережі довгої короткострокової пам'яті (LSTM) та мережі суміші густин (MDN).

На етапі оцінювання ефективності запропонованих рішень особливу увагу приділено дотриманню методологічної цілісності дослідження. Методологічною основою дослідження є модель соціально-адаптивної навігації, представлена в розділі 2. 4.

Процес навчання PPO-LSTM-MDN здійснювався із застосуванням стратегії Curriculum Learning представленої у розділі 2. 6. Використання CL дозволило забезпечити стабільність збіжності алгоритму в умовах високої розмірності простору станів. Послідовний перехід від статичних сценаріїв до взаємодії з

рухомими агентами сприяв ефективному формуванню стійкої політики керування AMP.

З метою верифікації безпосереднього впливу методу адаптивного формування функції винагороди проведено валідаційне тестування. Навчання агента за PPO-LSTM-MDN відбувалося за умови використання ідентичних гіперпараметрів порівняно з моделлю PPO-CL. Такий підхід передбачав встановлення однакових значень швидкості навчання, коефіцієнта дисконтування, розміру пакету даних та параметрів оптимізації градієнта для обох архітектур.

Забезпечення ідентичності конфігураційних параметрів дозволило ізолювати та об'єктивно оцінити внесок модуля прогнозування невизначеності у підвищення якості навігації. Такий підхід гарантує валідність порівняння, доводячи, що зростання метрик успішності та соціальної прийнятності є наслідком застосування запропонованого методу, а не результатом підбору гіперпараметрів навчання.

Для валідації розробленого методу проведено серію порівняльних експериментів. У якості методів для порівняльного аналізу обрано алгоритм PPO з використанням Curriculum Learning (PPO-CL), реактивний алгоритм модель соціальних сил (SFM) та алгоритм оптимального взаємного уникнення зіткнень (ORCA).

Для оцінювання ефективності запропонованого алгоритму використано тестові сценарії, детальний опис яких наведено у розділі 4. 3. Тестування було проведено в умовах поступового зростання динамічної складності –статичне середовище (Zone 1), середовища з відповідно одним та двома рухомими агентами (Zone 2, Zone 3), середовище з підвищеною щільністю статичних та динамічних перешкод (Full Map).

У статичному середовищі (Zone 1) запропонований метод PPO-LSTM-MDN продемонстрував 100% успішність, подолавши маршрут у середньому за 18,10 с при довжині шляху 4,68 м. Алгоритм PPO CL продемонстрував дещо нижчий результат – 96%. Водночас алгоритм ORCA виявився найменш ефективним у статичному середовищі, досягнувши лише 72% успішності через значну кількість зіткнень. Це підтверджує недолік локальних геометричних планувальників, які

оптимізовані для уникнення динамічних агентів, проте часто потрапляють у локальні мінімуми поблизу статичних об'єктів. Модель SFM досягла показника 92%, але витратила на виконання завдання значно більше часу 30,28 с.

У сценарії Zone 2 запропонована модель PPO-LSTM-MDN зберегла абсолютну успішність 100%, продемонструвавши високий рівень соціальної прийнятності ($SCS = 72,3\%$). Час виконання становив 22,14 с, а довжина шляху – 5,59 м. Варто відзначити, що алгоритм SFM також виконав завдання у 100% випадків, однак значно збільшилися час навігації (40,02 с) та шлях (7,96 м), зменшився узагальнений показник відповідності соціальним нормам ($SCS = 52,2\%$).

Зі збільшенням кількості динамічних агентів у сценарії Zone 3 переваги адаптивного формування винагороди стають більш очевидними. PPO-LSTM-MDN успішно завершив 100% епізодів, обравши оптимальні траєкторії, про які свідчить довжина шляху 4,82 м та час 19,03 с. AMP завчасно коригував власну швидкість, уникаючи необхідності різких маневрів у безпосередній близькості до агентів.

Для порівняння, алгоритм PPO-CL без урахування невизначеності знизив успішність до 96%, а алгоритм ORCA продемонстрував показник SCS на рівні 69,0%, що вказує на часте порушення проксемічних зон людей.

У четвертому сценарії (Full Map) зафіксовано зниження показників ефективності для всіх тестованих алгоритмів. Однак запропонований метод зберіг найвищий узагальнений показник відповідності соціальним нормам серед усіх алгоритмів $SCS = 77,6\%$. Алгоритм ORCA у цьому ж сценарії показав різке падіння ефективності до 76% через зіткнення зі статичними об'єктами. Незважаючи на успішне завершення 92% епізодів, модель SFM продемонструвала зниження навігаційної ефективності. Зафіксований час виконання (63,46 с) вказує на формування надто повільного та обережного руху, який є неприйнятним для інтеграції робота в активний соціальний потік.

Узагальнені результати за всіма сценаріями (Overall) підтверджують перевагу запропонованого методу. Алгоритм PPO-LSTM-MDN досяг найвищого загального показника успішності 97,0%, випередивши SFM (96,0%), PPO-CL

(95,0%) та ORCA (86,0%). При цьому середній час проходження маршруту для розробленої моделі становить 20,99 с, що вдвічі швидше за SFM (41,45 с) і швидше за PPO-CL (22,65 с).

За узагальненим показником відповідності соціальним нормам метод PPO-LSTM-MDN перевищує алгоритм SFM та PPO-CL. Це підтверджує гіпотезу дисертаційного дослідження – адаптивна функція винагороди, яка враховує прогноз невизначеності на основі LSTM-MDN, стимулює AMP обирати соціально-прийнятні траєкторії без надмірного сповільнення.

Зведені показники ефективності алгоритмів навігації наведено у таблиці 4. 4.

Таблиця 4.4

Зведені показники ефективності алгоритмів навігації

Показники ефективності	PPO-LSTM-MDN (Запропонований)	PPO CL	SFM	ORCA
Success Rate, %	97	95	96	86
Human Collisions, %	1	2	2	2
Static Collisions, %	2	3	0	12
Timeouts, %	0	0	2	0
SCS	73,7	72,3	56,5	73,0
Time, с	20,99	22,65	41,45	22,90
Path, м	5,36	5,68	7,35	4,85

Для аналізу взаємозв'язку між часовою ефективністю та рівнем комфорту для оточуючих людей побудовано графік розсіювання. Дана візуалізація відображає пошук компромісу між часом виконання навігаційної задачі та метрикою соціальної прийнятності SCS.

З огляду на специфіку визначених параметрів, цільова область значень розташована у верхньому лівому куті координатної площини, який відповідає мінімальним часовим витратам за умови максимального дотримання соціальних норм (рис. 4. 7).

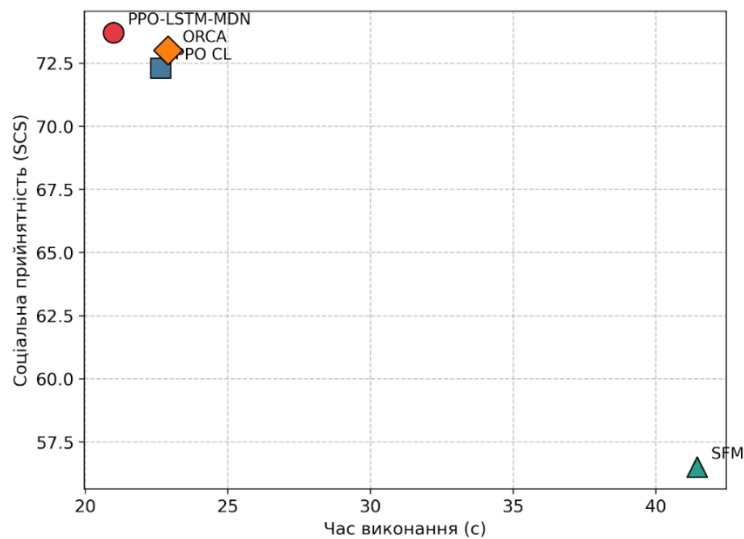


Рис. 4.7. Графік розсіювання показників часової ефективності та соціальної прийнятності досліджуваних алгоритмів

Аналіз розподілу маркерів дозволяє ідентифікувати перевагу запропонованого методу PPO-LSTM-MDN. Дана архітектура формує екстремум у безпосередній близькості до цільової зони, фіксуючи найменший середній час виконання (20,99 с) та найвищий показник соціальної прийнятності (73,7%).

Базовий алгоритм PPO-CL та реактивний планувальник ORCA утворюють суміжний кластер із близькими показниками. Незважаючи на близький рівень метрики SCS, обидва методи поступаються розробленій архітектурі за координатою часової ефективності.

Класична модель соціальних сил SFM демонструє найнижчу результативність, локалізуючись в правому нижньому квадранті графіка.

Отже, просторовий розподіл результатів верифікує спроможність моделі PPO-LSTM-MDN вирішувати фундаментальну проблему соціальної робототехніки – забезпечувати швидке просування цільовим маршрутом із зменшеною кількістю порушень проксемічних зон динамічних агентів.

Для комплексного оцінювання балансу характеристик алгоритмів побудовано пелюсткову діаграму (рис. 4. 8). Даний графік відображає п'ять ключових метрик: успішність (SUCC), соціальну прийнятність (SCS), соціальну безпеку (кількість зіткнень із людьми), а також ефективність часу та шляху. З метою порівняння параметрів, виражених у різних одиницях вимірювання,

застосовано процедуру нормалізації до 100-бальної шкали, де максимальне значення відповідає найкращому результату в межах вибірки.

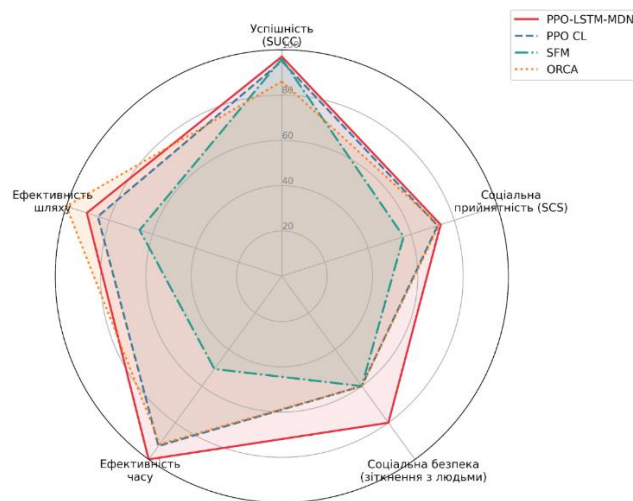


Рис. 4.8. Комплексне порівняння алгоритмів.

Аналіз побудованої діаграми підтверджує вибір найбільш розширеної та збалансованої сукупності показників у межах застосування методу PPO-LSTM-MDN (червона лінія). Зазначений алгоритм демонструє найвищий рівень соціальної безпеки. Даний показник розраховано на основі кількості контактів із людьми, де запропонований метод показав найкращий результат. Окрім того, модель PPO-LSTM-MDN забезпечує ефективність часу, випереджаючи ORCA, SFM та метод PPO-CL.

Порівняльний аналіз виявляє суттєву асиметрію в характеристиках інших методів. Алгоритм SFM (зелена лінія) показує прийнятну успішність, проте демонструє зниження значень за ефективністю часу та соціальною прийнятністю. Така конфігурація підтверджує схильність реактивних моделей до надмірної обережності, яка спричиняє сповільнення загального потоку руху. Натомість метод ORCA (помаранчева лінія) характеризується високою ефективністю пройденого шляху, але суттєво поступається за рівнем соціальної безпеки та успішності в умовах складних обмежень.

Підхід PPO CL (синя лінія) повторює загальну геометрію запропонованого методу, проте стабільно демонструє нижчі показники за кожною з осей. Це підтверджує наукову новизну та практичну цінність розробленої архітектури –

адаптивне формування функції винагороди на основі прогнозу невизначеності дозволяє досягти оптимального співвідношення між швидкістю виконання завдання та безпекою взаємодії в соціальному просторі.

Порівняльний аналіз PPO-LSTM-MDN із методом PPO-CL засвідчує комплексне покращення ключових навігаційних метрик завдяки впровадженню методу адаптивного формування функції винагороди. Інтеграція модуля прогнозування невизначеності дозволила знизити загальну кількість зіткнень, забезпечивши при цьому зменшення зіткнень із людьми. Окрім підвищення рівня соціальної безпеки, розроблена архітектура PPO-LSTM-MDN дозволила оптимізувати просторово-часові характеристики: середня довжина пройденого шляху скоротилася з 5,63%, а загальний час виконання зменшився на 7,33%. Відповідне зростання успішності епізодів та показника соціальної прийнятності підтвердило здатність агента проактивно реагувати на соціальну динаміку. Замість реактивного гальмування чи надмірної обережності, АМР здійснював завчасні коригування траєкторії руху, доводячи ефективність та доцільність запропонованої алгоритмічної модифікації.

На основі отриманих експериментальних даних можна стверджувати, що інтеграція моделі прогнозування невизначеності у процес формування функції винагороди алгоритму PPO забезпечила оптимальне співвідношення між безпекою, соціальною прийнятністю та швидкістю навігації.

Висновки до розділу 4

У четвертому розділі проведено експериментальні дослідження розроблених методів соціально-адаптивної навігації автономних мобільних роботів. На основі отриманих результатів сформульовано наступні висновки.

Розроблено та програмно реалізовано симуляційне середовище на базі інтерфейсу Gymnasium, інтегрований із моделлю соціальних сил (SFM). Архітектура середовища дозволяє здійснювати високопродуктивні ітераційні обчислення в 2D-просторі та забезпечити валідацію нейромережевих політик

керування. Ключовою особливістю реалізації є впровадження механізмів подолання розриву між симуляцією та реальністю. Це досягнуто шляхом математичного моделювання стохастичних похибок лазерного далекоміра та імітації латентності передачі даних через FIFO-буфери. Такий підхід забезпечує формування робастних стратегій керування, стійких до апаратних недосконалостей реальних робототехнічних систем.

Обґрунтовано та впроваджено багатокритеріальну систему оцінювання ефективності соціальної навігації. Окрім стандартних технічних метрик (коефіцієнт успішності, час та довжина шляху), запропоновано та математично формалізовано узагальнений показник відповідності соціальним нормам (Social Compliance Score, SCS). Даний показник дозволяє кількісно вимірювати рівень соціальної прийнятності руху АМР на основі аналізу часу порушення зон проксеміки та випадків фронтального зближення.

Експериментально підтверджено ефективність методу соціально-адаптивної навігації на основі PPO та Curriculum Learning (PPO-CL). Встановлено, що поетапне ускладнення навігаційних сценаріїв та автоматизований перехід між етапами на основі критеріїв стабільності навчання дозволяють пришвидшити навчання нейронної мережі. Застосування розробленого механізму відпалу соціального компонента функції винагороди дозволило уникнути нестабільності навчання на початкових фазах навчання на третьому етапі. Порівняльний аналіз продемонстрував, що метод PPO-CL забезпечує зростання успішності виконання завдань з 87% до 95% та зниження кількості критичних зіткнень із людьми порівняно із алгоритмом PPO.

Проведено верифікацію методу адаптивного формування винагороди на основі прогнозу невизначеності (PPO-LSTM-MDN). Завдяки використанню MDN для оцінки ймовірнісних характеристик майбутнього стану середовища, АМР отримав здатність проактивно коригувати швидкість і напрямок руху в умовах високої динамічної невизначеності. Результати валідації у соціальних сценаріях продемонстрували перевагу розробленого методу над базовим підходами. PPO-LSTM-MDN забезпечив найвищий показник успішності та узагальнений показник

відповідності соціальним нормам, демонструючи при цьому вдвічі менший час виконання завдання порівняно з моделлю соціальних сил.

Аналіз розподілу результатів у просторі метрик «Час – SCS» підтвердив, що запропонована архітектура PPO-LSTM-MDN формує оптимальну модель поведінки, яка локалізується в цільовій області мінімальних часових витрат при максимальному дотриманні соціальних норм. Це доводить наукову гіпотезу про те, що врахування невизначеності середовища дозволяє AMP ефективно виконувати завдання навігації в соціальному середовищі.

Результати проведених досліджень свідчать про повне виконання завдань розділу та підтверджують практичну придатність розроблених програмних та алгоритмічних рішень для впровадження в сучасні автономні інтелектуальні системи, які функціонують у динамічних середовищах із присутністю людей.

ВИСНОВКИ

У дисертаційній роботі розв'язано актуальну науково-практичну задачу підвищення ефективності та соціальної прийнятності навігації автономних мобільних роботів у динамічних середовищах шляхом розробки та впровадження методів глибинного навчання з підкріпленням, стратегій навчання за програмою (Curriculum Learning) та механізмів імовірнісного прогнозування.

Основні наукові та практичні результати дослідження полягають у наступному.

1. На основі системного аналізу сучасних підходів до автономної навігації встановлено, що класичні детерміновані алгоритми планування траєкторії не забезпечують необхідний рівень адаптивності в умовах спільного з людьми простору. Виявлено, що ключовою проблемою існуючих систем є ігнорування соціального контексту та стохастичної природи людського руху і створення ситуацій психологічного дискомфорту для оточуючих. Обґрунтовано необхідність переходу від парадигми простого уникнення перешкод до моделювання соціально-адаптивної взаємодії на основі методів глибинного навчання з підкріпленням.

2. Формалізовано модель соціально-адаптивної навігації автономного мобільного робота, в якій задачу керування представлено як марковський процес прийняття рішень (MDP) із розширеним простором станів. Запропонований підхід інтегрує кількісні параметри проксеміки за Е. Холлом безпосередньо у функцію винагороди. Це дозволило трансформувати неявні соціальні норми у чіткі математичні обмеження, забезпечуючи навчання агента діяти не лише безпечно, а й передбачувано для людей.

3. Розроблено метод навчання автономного мобільного робота на основі стратегії навчання за програмою (Curriculum Learning), який передбачає ієрархічну декомпозицію задачі навігації на чотири послідовні етапи:

1. Етап базової навігації у статичному середовищі для формування первинної політики руху.

2. Етап уникнення динамічних перешкод без урахування соціального контексту.
3. Етап інтеграції соціальних норм шляхом впровадження штрафів за порушення зон проксемики.
4. Етап фінальної оптимізації в умовах масштабованого динамічного середовища з підвищеною щільністю агентів-людей.

Встановлено, що використання автоматизованого механізму переходу між етапами забезпечує стабільну збіжність градієнта нейронної мережі.

4. Для забезпечення проактивної поведінки робота в умовах невизначеності динамічного середовища спроектовано та обґрунтовано архітектуру інтелектуального агента на основі алгоритму PPO, інтегрованого з модулем імовірнісного прогнозування LSTM-MDN. Використання мереж суміші густин (MDN) у поєднанні з рекурентними мережами дозволило апроксимувати мультимодальний розподіл імовірностей майбутніх положень людей. Це дало змогу кількісно оцінити міру невизначеності середовища та перейти від реактивного до проактивного планування маневрів автономного мобільного робота.

5. Розроблено механізм формування функції винагороди через впровадження динамічного зважування компонентів. Розроблений метод забезпечив адаптивне коригування пріоритетів між швидкістю досягнення цілі та дотриманням соціального комфорту залежно від обчисленого рівня невизначеності. У ситуаціях із високою невизначеністю руху людей робот надавав перевагу консервативній стратегії (зменшення швидкості, збільшення дистанції до людей) та мінімізував ризик виникнення небезпечних ситуацій.

6. Створено симуляційне навчальне середовище на базі інтерфейсу Gymnasium та моделі соціальних сил (SFM). Для подолання розриву між симуляцією та реальністю (Sim-to-Real) враховано стохастичні похибки сенсорних систем та латентність обчислювальних процесів. Це забезпечує робастність сформованих стратегій при їх перенесенні на реальні фізичні платформи.

Впроваджено систему метрик для оцінки якості соціальної навігації. Експериментально підтверджено, що запропонований метод PPO-CL підвищує успішність виконання завдань до 95 % порівняно з 87 % у базовому PPO, одночасно знижуючи кількість критичних зближень із агентами-людьми.

Результати валідації розробленого методу PPO-LSTM-MDN у соціальних сценаріях продемонстрували його перевагу над базовим підходами. Метод PPO-LSTM-MDN забезпечив найвищий показник успішності та узагальнений показник відповідності соціальним нормам.

Отримані результати мають практичне значення для проєктування сучасних програмно-апаратних рішень у сфері автономної робототехніки, інтелектуального керування та систем прийняття рішень. Запропоновані методи та програмні рішення можуть бути використані для подальшого розвитку систем автономної навігації, робототехнічних платформ та інтелектуальних систем керування в умовах динамічної взаємодії з людьми.

Результати дисертаційного дослідження повністю підтверджують поставлену мету та свідчать про виконання всіх визначених наукових завдань.

Перспективи майбутніх досліджень доцільно спрямувати на вивчення групової соціально-адаптивної навігації у надгустих потоках людей та розширення системи соціальних норм.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Lee, H., Enriquez, J. L., & Lee, G. (2022). Robotics 4.0: Challenges and Opportunities in the 4th Industrial Revolution. *Journal of Internet Services and Information Security*, 12(4), 39-55.
2. International Organization for Standardization. (2023). *Robotics – vocabulary* (ISO Standard No. 8373:2023).
3. Considine, D. M., & Considine, G. D. (1986). Robot technology fundamentals. In *Standard handbook of industrial automation* (pp. 262-320). Boston, MA: Springer US.
4. Singh, G., & Banga, V. K. (2022). Robots and its types for industrial applications. *Materials Today: Proceedings*, 60, 1779-1786. <https://doi.org/10.1016/j.matpr.2021.12.426>
5. Iida, F. (2007). *Autonomous Robots: From Biological Inspiration to Implementation and Control*. George A. Bekey. (2005, MIT Press.) Hardcover, 577 pages. ISBN 0262025787. *Artificial Life*, 13(4), 419–421. <https://doi.org/10.1162/artl.2007.13.4.419>
6. Zou, A. M., Hou, Z. G., Fu, S. Y., & Tan, M. (2006). Neural networks for mobile robot navigation: a survey. In *International symposium on neural networks* (pp. 1218-1226). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11760023_177
7. Rubio, F., Valero, F., & Llopis-Albert, C. (2019). A review of mobile robots: Concepts, methods, theoretical framework, and applications. *International Journal of Advanced Robotic Systems*, 16 (2), 172988141983959. <https://doi.org/10.1177/1729881419839596>
8. ІВАНЮК, О. І. (2020). Navigation of autonomous systems based on situation control with dynamic replanning. *Системи обробки інформації*, (3(162)), 44 – 51. <https://doi.org/10.30748/soi.2020.162.05>
9. Siegwart, R., Nourbakhsh, I. R., & Scaramuzza, D. (2011). *Introduction to Autonomous Mobile Robots*. MA, USA: MIT Press, p. 472.

10. Blanco, J.-L., Gonzalez, J., & Fernandez-Madrigal, J.-A. (2007). Mobile robot ego-motion estimation by proprioceptive sensor fusion. *Y 2007 9th International Symposium on Signal Processing and Its Applications (ISSPA)*. IEEE. <https://doi.org/10.1109/isspa.2007.4555413>
11. Gorostiza, E. M., Lázaro Galilea, J. L., Meca Meca, F. J., Salido Monzú, D., Espinosa Zapata, F., & Pallarés Puerto, L. (2011). Infrared Sensor System for Mobile-Robot Positioning in Intelligent Spaces. *Sensors*, 11(5), 5416-5438. <https://doi.org/10.3390/s110505416>
12. Aqel, M. O., Marhaban, M. H., Saripan, M. I., & Ismail, N. B. (2016). Review of visual odometry: types, approaches, challenges, and applications. *SpringerPlus*, 5(1), 1897.
13. Ганенко, Л. Д. (2023). Особливості планування шляху мобільного робота. *Технологійні горизонти: дослідження та застосування інформаційних технологій для технологійного прогресу України та світу: зб. тез всеукр. наук.-техн. конф.* (с. 193–195). Київ.
14. Barber, R., Crespo, J., Gómez, C., C. Hernández, A., & Galli, M. (2019). Mobile Robot Navigation in Indoor Environments: Geometric, Topological, and Semantic Navigation. *Y Applications of Mobile Robots*. IntechOpen. <https://doi.org/10.5772/intechopen.79842>
15. Elfes, A. (1989). Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6), 46-57. <https://doi.org/10.1109/2.30720>
16. Alatise, M. B., & Hancke, G. P. (2020). A review on challenges of autonomous mobile robot and sensor fusion methods. *IEEE access*, 8, 39830-39846. <https://doi.org/10.1109/ACCESS.2020.2975643>
17. Nüchter, A., & Hertzberg, J. (2008). Towards semantic maps for mobile robots. *Robotics and Autonomous Systems*, 56(11), 915-926. <https://doi.org/10.1016/j.robot.2008.08.001>
18. Ганенко, Л. Д. (2024). Моделювання середовища автономних мобільних роботів. *Технологічні горизонти: дослідження та застосування*

інформаційних технологій для технологічного прогресу України і світу: зб. тез II всеукр. наук.-техн. конф. (с. 89–90). Київ.

19. Kalman Filter and Its Application / Q. Li et al. 2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS), Tianjin, China, 1–3 November 2015. 2015. <https://doi.org/10.1109/icinis.2015.35> .

20. Fox, D., Burgard, W., & Thrun, S. (1999). Markov localization for mobile robots in dynamic environments. *Journal of artificial intelligence research*, 11, 391-427. <https://doi.org/10.1613/jair.616>

21. Dellaert, F., Fox, D., Burgard, W., & Thrun, S. (1999, May). Monte carlo localization for mobile robots. In *Proceedings 1999 IEEE international conference on robotics and automation (Cat. No. 99CH36288C)* (Vol. 2, pp. 1322-1328). IEEE. <https://doi.org/10.1109/ROBOT.1999.772544>

22. Ганенко, Л. Д. (2024). Інтелектуальні методи навігації мобільних роботів на основі SLAM: виклики та перспективи. *Цифрова гуманістика: Інформаційні технології та інформаційне моделювання на сучасному етапі розвитку суспільства: зб. тез всеукр. наук.-практ. конф.* (с. 186–189). Кропивницький.

23. Hanenko, L., Storchak, K., Shlianchak, S., Vorokhob, M., & Pitaichuk, M. (2025). SLAM in navigation systems of autonomous mobile robots. *Cybersecurity Providing in Information and Telecommunication Systems 2025*, (3991), 173-182. <https://ceur-ws.org/Vol-3991/paper13.pdf>

24. Jang, H., Kim, T., Ahn, K., Jeon, S., & Kang, Y. (2024). Dynamic Occupancy Grid Map with Semantic Information Using Deep Learning-Based BEVFusion Method with Camera and LiDAR Fusion. *Sensors*, 24(9), 2828. <https://doi.org/10.3390/s24092828>

25. Jiménez Schlegl, P., Thomas, F., & Torras, C. (2001). 3D collision detection: a survey. *Computers & Graphics*, Vol. 25, Issue 2, p. 269-285, [https://doi.org/10.1016/S0097-8493\(00\)00130-8](https://doi.org/10.1016/S0097-8493(00)00130-8)

26. Borenstein, J., & Koren, Y. (1991). The vector field histogram-fast obstacle avoidance for mobile robots. *IEEE Transactions on Robotics and Automation*, 7(3), 278–288. <https://doi.org/10.1109/70.88137>
27. Fox, D., Burgard, W., & Thrun, S. (1997). The dynamic window approach to collision avoidance. *IEEE Robotics & Automation Magazine*, 4(1), 23–33. <https://doi.org/10.1109/100.580977>
28. Simmons, R. (1996, April). The curvature-velocity method for local obstacle avoidance. In *Proceedings of IEEE international conference on robotics and automation* (Vol. 4, p. 3375-3382). IEEE. <https://doi.org/10.1109/ROBOT.1996.511023>
29. Ганенко, Л. Д., Жебка, В. В. (2023). Аналітичний огляд питань навігації мобільних роботів в закритих приміщеннях. *Телекомунікаційні та інформаційні технології*, (3), 85-96. <https://doi.org/10.31673/2412-4338.2023.038087>
30. Debnath, S., Omar, R., Bagchi, S., Sabudin, E. N., Shee Kandar, M. H. A., Foysol, K., & Chakraborty, T. (2021). Different cell decomposition path planning methods for unmanned air vehicles: A review. In *Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019* (pp. 99–111). Springer, Singapore. https://doi.org/10.1007/978-981-15-5281-6_8
31. El Khaili, M. (2014). Visibility graph for path planning in the presence of moving obstacles. *Engineering Science and Technology: An International Journal*, 4(2), 118–123.
32. Siegwart, R., Nourbakhsh, I. R., & Scaramuzza, D. (2011). *Introduction to Autonomous Mobile Robots*. MIT Press.
33. Latombe, J. C. (1991). *Robot motion planning*. Springer US.
34. Проценко, А. А., Иванов, В. Г. (2019). Класичні методи планування шляху для мобільних роботів. *Системи управління, навігації та зв'язку. Збірник наукових праць*. 3(55), 143–151. <https://doi.org/10.26906/sunz.2019.3.143>
35. Yahja, A., Stentz, A., Singh, S., & Brumitt, B. L. (1998, May). Framed-quadtree path planning for mobile robots operating in sparse environments. In *Proceedings. 1998 IEEE international conference on robotics and automation* (Cat. No. 98CH36146) (Vol. 1, pp. 650-655). IEEE. <https://doi.org/10.1109/robot.1998.677046>

36. Dobrin, A. (2005). A review of properties and variations of Voronoi diagrams. *Whitman College*, 10, 9156.
37. Kavraki, L., Svestka, P., Latombe, J.-C., & Overmars, M. H. (1996). Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Transactions on Robotics and Automation*, 12(4), 566–580. <https://doi.org/10.1109/70.508439>.
38. Alshammrei, S., Boubaker, S., & Kolsi, L. (2022). Improved Dijkstra algorithm for mobile robot path planning and obstacle avoidance. *Computers, Materials & Continua*, 72(3), 5939–5954. <https://doi.org/10.32604/cmc.2022.028165>.
39. Guruji, A. K., Agarwal, H., & Parsediya, D. K. (2016). Time-efficient A* algorithm for robot path planning. *Procedia Technology*, 23, 144–149. <https://doi.org/10.1016/j.protcy.2016.03.010>.
40. Guo, J., & Liu, L. (2010). A study of improvement of D* algorithm for mobile robot path planning in partial unknown environment. *Kybernetes*, 39(6), 935–945. <https://doi.org/10.1108/03684921011046708>.
41. Остапенко, Л., Харченко, В. (2024). Алгоритми формування маршрутів в 2d/3d просторі з використанням мобільних засобів для забезпечення комунікацій в умовах руйнувань. *Вимірювальна та обчислювальна техніка в технологічних процесах*, (2), 168–178. <https://doi.org/10.31891/2219-9365-2024-78-20>.
42. Павлюк, О. М., Медиковський, М. О. ., Міщук, М. В., & Заболотна, А. О. (2025). Метод визначення оптимального шляху мобільної роботизованої платформи в умовах обмежених ресурсів. *Вісник Вінницького політехнічного інституту*, (1), 7–17. <https://doi.org/10.31649/1997-9266-2025-178-1-7-17>.
43. Sabudin, E. N., Omar, R., & Hailma, C. K. N. (2016). Potential field methods and their inherent approaches for path planning. *Journal of Engineering and Applied Sciences*, 11(18), 10801–10805.
44. Берізка, І. А., Карбовник, І. Д. (2024). Математична модель модифікованого методу штучних потенціальних полів з використанням функції Лапласа для уникнення перешкод в режимі реального часу. *Прикладні проблеми*

- комп'ютерних наук, безпеки та математики, 3, 12-22.
<https://apcssm.vnu.edu.ua/index.php/Journalone/article/view/123>
45. Sampathkumar, S. K., Choi, D., & Kim, D. (2024). Fuzzy inference system-assisted human-aware navigation framework based on enhanced potential field. *Complex Engineering Systems*, 4, 3. <https://doi.org/10.20517/ces.2023.34>
46. Borenstein, J., & Koren, Y. (1989). Real-time obstacle avoidance for fast mobile robots. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5), 1179–1187. <https://doi.org/10.1109/21.44033>.
47. Ганенко, Л. Д. (2024). Методи уникнення рухомих перешкод автономним мобільним роботом. *Проблеми комп'ютерної інженерії: зб. тез V всеукр. наук.-практ. конф.* (с. 6–8). Київ.
48. Borenstein, J., & Koren, Y. (1991). The vector field histogram-fast obstacle avoidance for mobile robots. *IEEE Transactions on Robotics and Automation*, 7(3), 278–288. <https://doi.org/10.1109/70.88137>.
49. Чумак, О., Дудко, М., & Дмитрієв, О. (2024). Онтологія методів планування маршрутів руху безпілотних літальних апаратів. *Випробування та сертифікація*, (1(3), 69-77. <https://doi.org/10.37701/ts.03.2024.10>.
50. Berg, J., Snape, J., Guy, S. J., & Manocha, D. (2011b). Reciprocal collision avoidance with acceleration-velocity obstacles. *У 2011 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. <https://doi.org/10.1109/icra.2011.5980408>
51. Helbing, D., & Molnár, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, 51(5), 4282–4286. <https://doi.org/10.1103/PhysRevE.51.4282>
52. Truong, X. T., & Ngo, T. D. (2017). Toward socially aware robot navigation in dynamic and crowded environments: A proactive social motion model. *IEEE Transactions on Automation Science and Engineering*, 14(4), 1743-1760. <https://doi.org/10.1109/tase.2017.2731371>
53. Tang, Y., Zakaria, M. A., & Younas, M. (2025). Path planning trends for autonomous mobile robot navigation: A review. *Sensors*, 25(4), 1206. <https://doi.org/10.3390/s25041206>

54. Pradhan, S. K., Parhi, D. R., & Panda, A. K. (2009). Fuzzy logic techniques for navigation of several mobile robots. *Applied Soft Computing*, 9(1), 290–304. <https://doi.org/10.1016/j.asoc.2008.04.008>
55. Lamini, C., Benhlila, S., & Elbekri, A. (2018). Genetic algorithm based approach for autonomous mobile robot path planning. *Procedia Computer Science*, 127, 180–189. <https://doi.org/10.1016/j.procs.2018.01.113>
56. Zheng, L., Yu, W., Li, G., Qin, G., & Luo, Y. (2023). Particle Swarm Algorithm Path-Planning Method for Mobile Robots Based on Artificial Potential Fields. *Sensors*, 23(13), 6082. <https://doi.org/10.3390/s23136082>
57. Wu, L., Huang, X., Cui, J., Liu, C., & Xiao, W. (2023). Modified adaptive ant colony optimization algorithm and its application for solving path planning of mobile robot. *Expert Systems with Applications*, 215, Article 119253. <https://doi.org/10.1016/j.eswa.2022.119410>
58. Contreras-Cruz, M. A., Ayala-Ramirez, V., & Hernandez-Belmonte, U. H. (2015). Mobile robot path planning using artificial bee colony and evolutionary programming. *Applied Soft Computing*, 30, 319–328. <https://doi.org/10.1016/j.asoc.2015.01.067>
59. Patle, B. K., Pandey, A., Parhi, D. R., & Jagadeesh, A. (2019). A review: Path planning strategies for mobile robot navigation. *Defense Technology*, 15(4), 582–606. <https://doi.org/10.1016/j.dt.2019.04.011>
60. Chen, Y. F., Liu, M., Everett, M., & How, J. P. (2017). Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 285–292. <https://doi.org/10.1109/icra.2017.7989037>
61. Everett, M., Chen, Y. F., & How, J. P. (2021). Collision Avoidance in Pedestrian-Rich Environments With Deep Reinforcement Learning. *IEEE Access*, 9, 10357–10377. <https://doi.org/10.1109/access.2021.3050338>
62. Li, K., Xu, Y., Wang, J., & Meng, M. Q. H. (2019). SARL*: Deep Reinforcement Learning based Human-Aware Navigation for Mobile Robot in Indoor

Environments. У *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE. 688-694 <https://doi.org/10.1109/robio49542.2019.8961764>

63. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social LSTM: Human Trajectory Prediction in Crowded Spaces. У *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. <https://doi.org/10.1109/cvpr.2016.110>

64. Ганенко, Л. Д. (2025). Методи прогнозування руху людини в контексті безпечної взаємодії з мобільними роботами. *Цифрова гуманістика: Інформаційні технології та інформаційне моделювання на сучасному етапі розвитку суспільства: зб. тез всеукр. наук.-практ. конф.* (с. 208–210). Кропивницький.

65. Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., & Alahi, A. (2018). Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1345–1354. <https://doi.org/10.1109/cvpr.2018.00240>

66. Kosaraju, V., Sadeghian, A., Martín-Martín, R., Reid, I.D., RezaTofighi, S.H., & Savarese, S. (2019). Social-BiGAT: Multimodal Trajectory Forecasting using Bicycle-GAN and Graph Attention Networks. *ArXiv, abs/1907.03395*.

67. Singamaneni, P. T., Bachiller-Burgos, P., Manso, L. J., Garrell, A., Sanfeliu, A., Spalanzani, A., & Alami, R. (2024). A survey on socially aware robot navigation: Taxonomy and future challenges. *The International Journal of Robotics Research*, 43(10), 1533-1572. <https://doi.org/10.1177/02783649241230562>

68. Ngo, H. Q. T., Le, V. N., Thien, V. D. N., Nguyen, T. P., & Nguyen, H. (2020). Develop the socially human-aware navigation system using dynamic window approach and optimize cost function for autonomous medical robot. *Advances in Mechanical Engineering*, 12(12), 168781402097943. <https://doi.org/10.1177/1687814020979430>

69. Wang, Y., Yu, J., Kong, Y., Sun, L., Liu, C., Wang, J., & Chi, W. (2024). Socially Adaptive Path Planning Based on Generative Adversarial Network. *IEEE Transactions on Intelligent Vehicles*, 1–13. <https://doi.org/10.1109/tiv.2024.3478219>

70. Singamaneni, P. T., Bachiller-Burgos, P., Manso, L. J., Garrell, A., Sanfeliu, A., Spalanzani, A., & Alami, R. (2024). A survey on socially aware robot navigation: Taxonomy and future challenges. *The International Journal of Robotics Research*. <https://doi.org/10.1177/02783649241230562>
71. Li, K., Lu, Y., & Meng, M. Q. H. (2021). Human-Aware Robot Navigation via Reinforcement Learning with Hindsight Experience Replay and Curriculum Learning. 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE. <https://doi.org/10.1109/robio54168.2021.9739519>
72. Daza, M., Barrios-Aranibar, D., Diaz-Amado, J., Cardinale, Y., & Vilasboas, J. (2021). An Approach of Social Navigation Based on Proxemics for Crowded Environments of Humans and Robots. *Micromachines*, 12(2), 193. <https://doi.org/10.3390/mi12020193>
73. Kang S, Yang S, Kwak D, Jargalbaatar Y, Kim D. Social Type-Aware Navigation Framework for Mobile Robots in Human-Shared Environments. *Sensors*. 2024; 24(15):4862. <https://doi.org/10.3390/s24154862>
74. Ганенко, Л. Д., В. В. Жебка. Застосування методів навчання з підкріпленням для планування шляху мобільних роботів. Телекомунікаційні та інформаційні технології 1 (2024): 16-25. <https://doi.org/10.31673/2412-4338.2024.011625>.
75. Sun, H., Zhang, W., Yu, R., & Zhang, Y. (2021). Motion planning for mobile robots—Focusing on deep reinforcement learning: A systematic review. *IEEE Access*, 9, 69061-69081. <https://doi.org/10.1109/ACCESS.2021.3076530>.
76. Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11), 1238-1274. <https://doi.org/10.1177/0278364913495721>
77. Hu, C., Ning, B., Xu, M., & Gu, Q. (2020). An experience aggregative reinforcement learning with multi-attribute decision-making for obstacle avoidance of wheeled mobile robot. *IEEE Access*, 8, 108179-108190. doi: 10.1109/ACCESS.2020.3001143

78. Zhu, K., & Zhang, T. (2021). Deep reinforcement learning based mobile robot navigation: A review. *Tsinghua Science and Technology*, 26(5), 674-691. doi: 10.26599/TST.2021.9010012
79. Kommey, B., Isaac, O. J., Tamakloe, E., & Opoku, D. (2024). A Reinforcement Learning Review: Past Acts, Present Facts and Future Prospects. *IT Journal Research and Development*, 8(2), 120-142. <https://doi.org/10.25299/itjrd.2023.13474>
80. Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019). Q-learning algorithms: A comprehensive classification and applications. *IEEE access*, 7, 133653-133667. <https://doi.org/10.1109/ACCESS.2019.2941229>
81. Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019). Q-learning algorithms: A comprehensive classification and applications. *IEEE access*, 7, 133653-133667. <https://doi.org/10.1109/ACCESS.2019.2941229>
82. Mitić, M., Miljković, Z., & Babić, B. (2011). Empirical control system development for intelligent mobile robot based on the elements of the reinforcement machine learning and axiomatic design theory. *FME Transactions, New Series*, 39(1), 1-8.
83. Zhao, Y., Zhang, Y., & Wang, S. (2021, December). A review of mobile robot path planning based on deep reinforcement learning algorithm. In *Journal of Physics: Conference Series* (Vol. 2138, No. 1, p. 012011). IOP Publishing. <https://doi.org/10.1088/1742-6596/2138/1/012011>
84. Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3-4), 229–256. <https://doi.org/10.1007/BF00992696>
85. Roderick, M., MacGlashan, J., & Tellex, S. (2017). Implementing the deep q-network. *arXiv preprint arXiv:1711.07478*. <https://doi.org/10.48550/arXiv.1711.07478>
86. Jafari, S., Hoseinzadeh, S., & Sohani, A. (2022). Deep Q-value neural network (DQN) reinforcement learning for the techno-economic optimization of a solar-

driven nanofluid-assisted desalination technology. *Water*, 14(14), 2254. <https://doi.org/10.3390/w14142254>

87. Yang, Y., Juntao, L., & Lingling, P. (2020). Multi-robot path planning based on a deep reinforcement learning DQN algorithm. *CAAI Transactions on Intelligence Technology*, 5(3), 177-183. <https://doi.org/10.1049/trit.2020.0024>

88. Van Hasselt H., Guez A., Silver D. Deep Reinforcement Learning with Double Q-Learning. Proceedings of the AAAI Conference on Artificial Intelligence. 2016. Vol. 30, no. 1. URL: <https://doi.org/10.1609/aaai.v30i1.10295>

89. Quinones-Ramirez, M., Rios-Martinez, J., & Uc-Cetina, V. (2023). Robot path planning using deep reinforcement learning. *arXiv preprint arXiv:2302.09120*. <https://doi.org/10.48550/arXiv.2302.09120>

90. Millan-Arias, C. C., Fernandes, B. J., Cruz, F., Dazeley, R., & Fernandes, S. (2021). A robust approach for continuous interactive actor-critic algorithms. *Ieee Access*, 9, 104242-104260. <https://doi.org/10.1109/ACCESS.2021.3099071>

91. Barto, A. G., Sutton, R. S., & Anderson, C. W. (2020). Looking back on the actor–critic architecture. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(1), 40-50. <https://doi.org/10.1109/TSMC.2020.3041775>

92. Gu, Z., Jia, Z., & Choset, H. (2019). Adversary a3c for robust reinforcement learning. *arXiv preprint arXiv:1912.00330*. <https://doi.org/10.48550/arXiv.1912.00330>

93. Wang, H., Gao, W., Wang, Z., Zhang, K., Ren, J., Deng, L., & He, S. (2023). Research on obstacle avoidance planning for UUV based on A3C algorithm. *Journal of Marine Science and Engineering*, 12(1), 63. <https://doi.org/10.3390/jmse12010063>

94. Ганенко, Л. Д. (2024). Методи навчання з підкріпленням у навігаційних системах мобільних роботів. *Інновації: наук. конф. молодих вчен.* (с. 28–30). Київ.

95. Aydogmus, O., & Yilmaz, M. (2023). Comparative analysis of reinforcement learning algorithms for bipedal robot locomotion. *IEEE Access*, 12, 7490-7499. <https://doi.org/10.1109/ACCESS.2023.3344393>

96. Zhao, T., Wang, M., Zhao, Q., Zheng, X., & Gao, H. (2023). A path-planning method based on improved soft actor-critic algorithm for mobile robots. *Biomimetics*, 8(6), 481. <https://doi.org/10.3390/biomimetics8060481>

97. Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018, July). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning* (pp. 1861-1870).
98. Chen, Y., & Liang, L. (2023). SLP-improved DDPG path-planning algorithm for mobile robot in large-scale dynamic environment. *Sensors*, 23(7), 3521. <https://doi.org/10.3390/s23073521>
99. Li, P., Ding, X., Sun, H., Zhao, S., & Cajo, R. (2021). Research on dynamic path planning of mobile robot based on improved DDPG algorithm. *Mobile Information Systems*, 2021(1), 5169460. <https://doi.org/10.1155/2021/5169460>
100. Gong, H., Wang, P., Ni, C., & Cheng, N. (2022). Efficient path planning for mobile robot based on deep deterministic policy gradient. *Sensors*, 22(9), 3579. <https://doi.org/10.3390/s22093579>
101. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*. <https://doi.org/10.48550/arXiv.1707.06347>
102. International Organization for Standardization. (2014). *Robots and robotic devices – Safety requirements for personal care robots* (ISO Standard No. 13482:2014). <https://www.iso.org/standard/53820.html>
103. Ганенко, Л. Д. (2025). Застосування DRL для соціальної навігації автономного мобільного робота. *Проблеми комп'ютерної інженерії: зб. тез VI Всеукр. наук.-практ. конф.* Київ.
104. Ганенко, Л., Жебка, В. (2025). Модель соціально-адаптивної навігації мобільного робота з використанням методів навчання з підкріпленням. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*, 1(29), 559-570. <https://doi.org/10.28925/2663-4023.2025.29.907>
105. Thrun, S., Burgard, W., & Fox, D. (2005). *Probabilistic Robotics*. MIT Press.
106. Soviany, Petru, et al. Curriculum learning: A survey. *International Journal of Computer Vision* 130.6 (2022): 1526-1565. <https://doi.org/10.1007/s11263-022-01611-x>.

107. Ганенко, Л. Д., Жебка, В. В. (2025). Curriculum learning як стратегія оптимізації навчання робототехнічних систем. *Виклики та рішення в програмній інженерії: зб. тез Всеукр. наук.-тех. конф.* (с. 366–368). Київ.
108. Wang, X., Chen, Y., & Zhu, W. (2021). A Survey on Curriculum Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1, <https://doi.org/10.1109/tpami.2021.3069908>.
109. Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. *Proceedings of the 26th Annual International Conference on Machine Learning*, 41–48. <https://doi.org/10.1145/1553374.1553380>.
110. Ганенко, Л., Бушма, О. (2025). Метод навчання автономних мобільних роботів на основі DRL та Curriculum Learning. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*, 2(30), 568–582. <https://doi.org/10.28925/2663-4023.2025.30.994>.
111. Anca, M., Thomas, J. D., Pedamonti, D., Hansen, M., & Studley, M. (2023, October). Achieving goals using reward shaping and curriculum learning. In *Proceedings of the Future Technologies Conference* (pp. 316–331). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-47454-5_24.
112. Narvekar, S., Peng, B., Leonetti, M., Sinapov, J., Taylor, M. E., & Stone, P. (2020). Curriculum learning for reinforcement learning domains: A framework and survey. *Journal of Machine Learning Research*, 21(181), 1–50. <https://doi.org/10.48550/arXiv.2003.04960>
113. Florensa, C., Held, D., Wulfmeier, M., Zhang, M., & Abbeel, P. (2018). Reverse curriculum generation for reinforcement learning agents. *Proceedings of the 31st Conference on Learning Theory*, 75, 482–495. <https://doi.org/10.48550/arXiv.1707.05300>
114. Ng, A. Y., Harada, D., & Russell, S. (1999). Policy invariance under reward shaping: Theory and application to predicting reinforcement learning. *Proceedings of the 16th International Conference on Machine Learning (ICML 99)*, 278–287.

115. Freitag, K., Ceder, K., Laezza, R., Åkesson, K., & Chehreghani, M. H. (2024). Curriculum Reinforcement Learning for Complex Reward Functions. *arXiv preprint arXiv:2410.16790*. <https://doi.org/10.48550/arXiv.2410.16790>
116. Ганенко, Л. Д. (2025). Метод соціально-адаптивної навігації мобільного робота. *Сучасні аспекти діджиталізації та інформатизації в програмній та комп'ютерній інженерії: зб. тез III Міжнар. конф.* Київ.
117. Xiang, W., YIN, H., Wang, H., & Jin, X. (2024). SocialCVAE: Predicting Pedestrian Trajectory via Interaction Conditioned Latents. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(6), 6216–6224. <https://doi.org/10.1609/aaai.v38i6.28439>
118. Kosaraju, V., Sadeghian, A., Martín-Martín, R., Reid, I., Rezatofighi, H., & Savarese, S. (2019). Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks. *Advances in neural information processing systems*, 32. <https://proceedings.neurips.cc/paper/2019/file/d09bf41544a3365a46c9077ebb5e35c3-Paper.pdf>
119. Choi, S., Lee, K., Lim, S., & Oh, S. (2018). Uncertainty-aware learning from demonstration using mixture density networks with sampling-free variance modeling. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 6915–6922. <https://doi.org/10.48550/arXiv.1709.02249>
120. Bishop, C. M. (1994). Mixture density networks (Technical Report NCRG/4300). Aston University.
121. Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*. <https://doi.org/10.48550/arXiv.1308.0850>
122. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
123. Devidze, R. (2025). Reward design for reinforcement learning agents. *arXiv preprint arXiv:2503.21949*. <https://doi.org/10.48550/arXiv.2503.21949>
124. Ганенко Л. Д. (2026) Метод адаптивного формування винагороди за умов невизначеності динамічних об'єктів. *Телекомунікаційні та інформаційні технології*, 1, 23-30. <https://doi.org/10.31673/2412-4338.2026.019003>

125. Mengacci, R., Zambella, G., Grioli, G., Caporale, D., Catalano, M. G., & Bicchi, A. (2021). An open-source ROS-Gazebo toolbox for simulating robots with compliant actuators. *Frontiers in Robotics and AI*, 8, Article 713083. <https://doi.org/10.3389/frobt.2021.713083>
126. Varma, A. (2021). Simulation of indoor localization and navigation of Turtlebot 3 using real time object detection. *2021 IEEE International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, 222-227. <https://doi.org/10.1109/CENTCON52345.2021.9687937>.
127. Гуржій С. В. (2025). ROS 2: особливості архітектури та практичного впровадження. *Вчені записки ТНУ імені В.І. Вернадського. Серія: Технічні науки*. 36 (75), 157-164. <https://doi.org/10.32782/2663-5941/2025.3.2/22>.
128. Ганенко, Л. Д. (2024). Застосування ROS для розробки робототехнічних систем. *Сучасні аспекти діджиталізації та інформатизації в програмній та комп'ютерній інженерії: зб. тез II міжнар. конф.* (с. 151–153). Київ.
129. Ганенко, Л., Жебка, В. (2025). Створення навігаційної системи автономного мобільного робота засобами ROS 2. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*, 4(28), 498-510. <https://doi.org/10.28925/2663-4023.2025.28.824>.
130. Helbing, D., & Molnár, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, 51(5), 4282–4286. <https://doi.org/10.1103/PhysRevE.51.4282>
131. Mavrogiannis, C., Baldini, F., Wang, A., Zhao, D., Trautman, P., Steinfeld, A., Oh, J., & Schaal, S. (2023). Core challenges of social navigation: A survey. *ACM Transactions on Human-Robot Interaction*. <https://doi.org/10.1145/3583741>.

ДОДАТОК А. Список публікацій здобувача

Наукові праці, у яких опубліковані основні наукові результати дисертації:

1. Ганенко Л. Д., Жебка В. В. Аналітичний огляд питань навігації мобільних роботів в закритих приміщеннях. *Телекомунікаційні та інформаційні технології*. 2023. №3. С. 85-96.
2. Ганенко Л. Д., Жебка В. В. Застосування методів навчання з підкріпленням для планування шляху мобільних роботів. *Телекомунікаційні та інформаційні технології*. 2024. №1. С. 16-25.
3. Ганенко Л. Д., Жебка В. В. Створення навігаційної системи автономного мобільного робота засобами ROS 2. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*. 2025 №4(28), С. 498–510.
4. Ганенко Л. Д., Жебка В. В. Модель соціально-адаптивної навігації мобільного робота з використанням методів навчання з підкріпленням. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*. 2025 №1 (29), С. 559-570.
5. Ганенко Л. Д., Бушма О. В. Метод навчання автономних мобільних роботів на основі DRL та Curriculum Learning. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*. 2025. № 2(30), С. 568–582.
6. Ганенко Л. Д. Метод адаптивного формування винагороди за умов невизначеності динамічних об'єктів. *Телекомунікаційні та інформаційні технології*. 2026. №1. С. 23-30.

Наукові праці, які засвідчують апробацію матеріалів дисертації:

7. Ганенко Л. Д., Жебка В. В. Особливості планування шляху мобільного робота. *Технологічні горизонти: дослідження та застосування інформаційних технологій для технологічного прогресу України та світу: зб. тез Всеукр. наук.-техн. конф., м. Київ, 28 листопада 2023 р. Київ: ДУІКТ, 2023. С. 193-195.*

8. Ганенко Л. Д. Інтелектуальні методи навігації мобільних роботів на основі SLAM: виклики та перспективи. *Цифрова гуманістика: Інформаційні технології та інформаційне моделювання на сучасному етапі розвитку суспільства*: зб. тез Всеукр. наук.-практ. конф., м. Кропивницький, 4-5 червня 2024 р. Кропивницький: ЦДУ ім. В. Винниченка, 2024 С. 186-189
9. Ганенко Л. Д., Жебка В. В. Методи навчання з підкріпленням у навігаційних системах мобільних роботів. *Інновації*: тези доп. наук. конф. молодих вчен., м. Київ, 19 вересня 2024 р. Київ: ДУІКТ, 2024. С. 28–30.
10. Ганенко Л. Д., Жебка В. В. Моделювання середовища автономних мобільних роботів. *Технологічні горизонти: дослідження та застосування інформаційних технологій для технологічного прогресу України і світу*: зб. тез II Всеукр. наук.-техн. конф., м. Київ, 18 листопада 2024 р. Київ: ДУІКТ, 2024. С. 89–90.
11. Ганенко Л. Д. Методи уникнення рухомих перешкод автономним мобільним роботом. *Проблеми комп'ютерної інженерії*: зб. тез V Всеукр. наук.-практ. конф., м. Київ, 03 грудня 2024 р. Київ: ДУІКТ, 2024. С.6–8.
12. Ганенко Л. Д., Жебка В. В. Застосування ROS для розробки робототехнічних систем. *Сучасні аспекти діджиталізації та інформатизації в програмній та комп'ютерній інженерії*: зб. тез II Міжнар. конф., м. Київ, 19-21 грудня 2024 р. Київ: ДУІКТ, 2024. С.151-153.
13. Hanenko, L., Storchak, K., Shlianchak, S., Vorokhob, M., & Pitaichuk, M. SLAM in navigation systems of autonomous mobile robots. *Cybersecurity Providing in Information and Telecommunication Systems 2025: workshop proc.*, Kyiv, 2025. Vol. 3991, P.173–182. *Scopus*
14. Ганенко Л. Д. Методи прогнозування руху людини в контексті безпечної взаємодії з мобільними роботами *Цифрова гуманістика: Інформаційні технології та інформаційне моделювання на сучасному етапі розвитку суспільства*: зб. тез Всеукр. наук.-практ. конф., м. Кропивницький, 22-23 травня 2025 р. Кропивницький: ЦДУ ім. В. Винниченка, 2025. С. 208–210.

15. Ганенко Л. Д., Жебка В. В. Curriculum learning як стратегія оптимізації навчання робототехнічних систем. *Виклики та рішення в програмній інженерії*: зб. тез Всеукр. наук.-тех. конф., м. Київ, 26 листопада 2025 р. Київ: ДУІКТ, 2025. С. 366–368.

16. Ганенко Л. Д., Жебка В. В. Застосування DRL для соціальної навігації автономного мобільного робота. *Проблеми комп'ютерної інженерії*: зб. тез VI Всеукр. наук.-практ. конф., м. Київ, 03 грудня 2025 р. Київ: ДУІКТ, 2025 С. 176–178.

17. Ганенко Л. Д. Метод соціально-адаптивної навігації мобільного робота. *Сучасні аспекти діджиталізації та інформатизації в програмній та комп'ютерній інженерії*: зб. тез III Міжнар. конф., м. Київ, 4-6 грудня 2025 р. Київ: ДУІКТ, 2025. С.282–284.

ДОДАТОК Б. Акти впровадження

КИЇВСЬКИЙ СТОЛИЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ БОРИСА ГРІНЧЕНКА



BORYS GRINCHENKO
KYIV METROPOLITAN UNIVERSITY

ФАКУЛЬТЕТ
ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
ТА МАТЕМАТИКИ
вул. Левка Лук'яненка, 13-Б, м. Київ, Україна, 04207
Тел.: +380 44 428-34-14
fitm.kubg.edu.ua, fitm@kubg.edu.ua

FACULTY
OF INFORMATION TECHNOLOGIES
AND MATHEMATICS
13-B Levka Lukianenka St, Kyiv, Ukraine, 04207
Tel.: +380 44 428-34-14
fitm.kubg.edu.ua, fitm@kubg.edu.ua

21.04.2026 № 19/2

АКТ

про впровадження результатів дисертаційного дослідження
Ганенко Людмили Дмитрівни
на тему «Методи та модель інтелектуальної навігації автономних
мобільних роботів у динамічному середовищі на основі глибинного
навчання з підкріпленням»,
поданої на здобуття наукового ступеня доктора філософії
зі спеціальності 123 Комп'ютерна інженерія

Цим Актом, ґрунтуючись на рішенні кафедри інформаційної та кібернетичної безпеки імені професора Володимира Бурячка Факультету інформаційних технологій та математики Київського столичного університету імені Бориса Грінченка, засвідчуємо, що нижчеперелічені наукові положення, а саме:

1. Удосконалено модель соціально-адаптивної навігації автономного мобільного робота на основі формалізації марковського процесу прийняття рішень, яка, на відміну від існуючих, завдяки інтеграції кінематичних обмежень робота з параметрами соціальної взаємодії, дозволяє системі керування ідентифікувати та класифікувати потенційні проксемічні конфлікти в режимі реального часу й підвищити безпеку руху в динамічному соціальному середовищі.
2. Вперше розроблено метод навчання навігаційної політики на основі глибинного навчання з підкріпленням, який за рахунок застосування комбінованої стратегії Curriculum Learning із механізмом автоматизованого переходу між етапами складності дозволяє вирішити проблему розрідженої

винагороди та забезпечити прискорення збіжності нейромережевої моделі в умовах соціальної навігації.

3. Вперше розроблено метод адаптивного формування винагороди на основі прогнозу невизначеності динамічного середовища, який за рахунок впровадження механізму динамічного зважування компонентів функції винагороди дозволяє оптимізувати стратегію поведінки автономного мобільного робота залежно від рівня невизначеності середовища.

Розроблені особисто Ганенко Людмилою Дмитрівною у ході проведення нею дисертаційних досліджень та отримали високу оцінку при обговоренні на засіданнях кафедри інформаційної та кібернетичної безпеки імені професора Володимира Бурячка Факультету інформаційних технологій та математики Київського столичного університету імені Бориса Грінченка.

Зазначені наукові результати:

по-перше, впроваджені в освітній процес кафедри інформаційної та кібернетичної безпеки імені професора Володимира Бурячка факультету інформаційних технологій та математики Київського столичного університету імені Бориса Грінченка у робочих програмах навчальних дисциплін спеціальності 123 Комп'ютерна інженерія;

по-друге, впроваджені в програмно-апаратне забезпечення лабораторії вбудованих систем і 3D моделювання та центру моделювання та програмування.

Дослідження Ганенко Людмили Дмитрівни відповідає всім вимогам до організації наукового пошуку та дає позитивний результат у практичному застосуванні.

Декан

Факультету інформаційних технологій та математики
кандидат фізико-математичних наук,
старший науковий співробітник



Оксана ЛИТВИН

Затверджую

Перший проректор
Державного університету
інформаційно-комунікаційних
технологій

Корченко О. Г.
"07" 09 2026 р.

АКТ

впровадження наукових результатів дисертації
аспіранта Державного університету інформаційно-комунікаційних
технологій

Ганенко Людмили Дмитрівни

Комісія у складі голови комісії – директора навчально-наукового інституту Інформаційних технологій доктора технічних наук, професора Нестеренко К.С. та членів комісії: завідувача кафедри Комп'ютерної інженерії кандидата технічних наук, доцента Лашевської Н.О., доцента кафедри Технологій цифрового розвитку кандидата фізико-математичних наук, доцента Поперешняк С.В. розглянула запропоновані матеріали в межах дисертаційного дослідження аспіранта Ганенко Людмили Дмитрівни на тему “Методи та модель інтелектуальної навігації автономних мобільних роботів у динамічному середовищі на основі глибинного навчання з підкріпленням”.

Комісією підтверджено, що реалізація поставленого наукового завдання сприяє підвищенню рівня безпеки та соціальної адаптивності автономних мобільних роботів у динамічних середовищах.

Під час виконання роботи створено та експериментально валідовано метод навчання навігаційної політики на основі глибинного навчання з


підкріпленням та Curriculum Learning. Застосування стратегії ітераційної модифікації середовища та поетапного ускладнення функції винагороди дозволило збільшити частку успішних епізодів з 87% до 95%.

На основі розробленого комплексного гібридного методу адаптивного формування функції винагороди PPO-LSTM-MDN для подальшого покращення взаємодії застосовано модуль оцінки невизначеності на основі рекурентних мереж. Результативність даного методу випередила результативність базових методів. При цьому загальний рівень успішного проходження маршруту зріс до 97%. Впровадження даного підходу дозволило оптимізувати просторово-часові характеристики руху автономного мобільного робота та збільшити показник відповідності соціальним нормам.


Практичне значення отриманих результатів полягає у можливості створення надійних систем керування інтелектуальними платформами. Отримані алгоритмічні рішення придатні для вдосконалення програмного забезпечення сервісної та промислової робототехніки з метою гарантування безпечного співіснування роботизованих систем та людей.

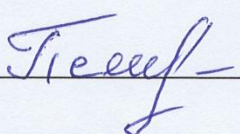
Теоретичні та практичні результати дисертаційного дослідження впроваджено в навчальний процес Навчально-наукового інституту інформаційних технологій Державного університету інформаційно-комунікаційних технологій.

Голова комісії:

 Катерина НЕСТЕРЕНКО

Члени комісії

 Наталія ЛАЩЕВСЬКА

 Світлани ПОПЕРЕШНЯК

АКТ

про впровадження результатів дисертаційної роботи

«16» серпня 2026 р.

м.Київ

Результати дисертаційної роботи Ганенко Людмили Дмитрівни на тему: «Методи та модель інтелектуальної навігації автономних мобільних роботів у динамічному середовищі на основі глибинного навчання з підкріпленням» були використані в діяльності компанії Byte Orchard Consulting у процесі дослідження та розробки програмних компонентів для систем автономної навігації та інтелектуального керування мобільними робототехнічними платформами. У межах впровадження застосовано модель соціально-адаптивної навігації на основі марковського процесу прийняття рішень, метод навчання навігаційної політики із використанням глибинного навчання з підкріпленням та Curriculum Learning, а також метод адаптивного формування функції винагороди на основі прогнозування невизначеності динамічного середовища.

Практичне використання результатів дисертаційного дослідження дозволило підвищити ефективність моделювання поведінки автономних агентів у динамічному середовищі, покращити адаптивність алгоритмів прийняття рішень в умовах невизначеності, а також забезпечити стабільність процесу навчання моделей глибинного навчання з підкріпленням. Запропоновані методи та програмні рішення можуть бути використані для подальшого розвитку систем автономної навігації, робототехнічних платформ та інтелектуальних систем керування в умовах динамічної взаємодії з людьми та рухомими об'єктами.

Отримані результати мають практичне значення для проектування сучасних програмно-апаратних рішень у сфері автономної робототехніки, інтелектуального керування та систем підтримки прийняття рішень.

Директор
ТОВ «Байт Орчард Консалтинг»



В. В. Рудич