УДК 004.896:623.746.7:355.41 DOI https://doi.org/10.35546/kntu2078-4481.2025.3.2.31

М. М. КОПИЛЕЦЬ
аспірант кафедри інформаційних систем на мереж
Національний університет «Львівська політехніка»
ORCID: 0009-0004-5823-9871

А. Я. САВКА
аспірант кафедри інформаційних систем на мереж
Національний університет «Львівська політехніка»
ORCID: 0009-0004-9198-6713

МЕТОДИ НАВЧАННЯ З ПІДКРІПЛЕННЯМ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ У ЗАВДАННЯХ ВІЙСЬКОВОЇ ЛОГІСТИКИ

У статті проведено розгорнутий огляд сучасних методів навчання з підкріпленням (Reinforcement Learning, RL) та їхнього застосування у сфері військової логістики із використанням безпілотних літальних апаратів (БпЛА). Актуальність теми зумовлена зростанням ролі БпЛА у забезпеченні оперативного транспортування вантажів, розвідки та підтримки бойових підрозділів, особливо в умовах обмеженого часу та високого рівня ризику. Проаналізовано ключові наукові роботи, що демонструють ефективність RL у завданнях планування маршруту, координації груп дронів (multi-agent RL, MARL), управління ресурсами та енергоспоживанням, а також у контексті врахування невизначеностей і ризиків (risk-sensitive RL, CVaR). Особливу увагу приділено підходам, які дозволяють формалізувати логістичні задачі як марковські процеси прийняття рішень із обмеженнями (CMDP), застосуванню механізмів уваги та графових нейронних мереж для оптимізації маршрутів, а також технологіям центрального навчання з децентралізованим виконанням (CTDE), що забезпечують ефективну взаємодію кількох БпЛА в реальному часі. Наведено математичні моделі та формули, що описують процес оптимізації політики керування, енергетичні обмеження та алгоритмічні модифікації, спрямовані на підвищення безпеки й надійності виконання завдань. Огляд містить аналіз підходів до інтеграції RL-рішень із системами моніторингу й контролю, а також описує сучасні виклики, серед яких є проблема перенесення навчених політик із симуляторів у реальні умови (sim-to-real), обмеженість обчислювальних ресурсів на борту БпЛА та необхідність стійкості до втрати зв'язку. Результати роботи можуть бути використані для побудови ефективних логістичних платформ, що здатні до автономної адаптації в умовах динамічних і небезпечних середовищ.

Ключові слова: *навчання з підкріпленням, безпілотні літальні апарати, військова логістика, MARL, CMDP, CVaR, CTDE, маршрутизація, енергетичні обмеження, симуляційне навчання.*

M. M. KOPYLETS
Postgraduate Student at the Department of Information Systems and Networks
Lviv Polytechnic National University
ORCID: 0009-0004-5823-9871

A. YA. SAVKA
Postgraduate Student at the Department of Information Systems and Networks
Lviv Polytechnic National University
ORCID: 0009-0004-9198-6713

REINFORCEMENT LEARNING METHODS FOR UNMANNED AERIAL VEHICLES IN MILITARY LOGISTICS TASKS

This article provides a comprehensive review of modern reinforcement learning (RL) methods and their application in military logistics involving unmanned aerial vehicles (UAVs). The relevance of this topic arises from the increasing role of UAVs in ensuring rapid cargo transportation, reconnaissance, and support for combat units, particularly under time constraints and high-risk conditions. The paper analyzes key peer-reviewed studies that demonstrate the efficiency of RL in route planning, multi-agent coordination (multi-agent RL, MARL), resource and energy management, and risk-sensitive decision-making (CVaR). Special attention is devoted to approaches that formalize logistics tasks as constrained Markov decision processes (CMDP), the application of attention mechanisms and graph neural networks for route optimization, and centralized training with decentralized execution (CTDE) frameworks that enable effective multi-UAV cooperation in real time. Mathematical models and formulas describing policy optimization, energy constraints, and algorithmic safety enhancements are presented. The review also examines methods for integrating RL solutions with monitoring and

control systems, addressing current challenges such as sim-to-real transfer, limited onboard computational resources, and resilience to communication loss. The findings of this work can serve as a foundation for building autonomous, adaptive logistics platforms capable of operating efficiently in dynamic and hazardous environments.

Key words: *reinforcement learning, unmanned aerial vehicles, military logistics, MARL, CMDP, CVaR, CTDE, routing, energy constraints, simulation-based training.*

Постановка проблеми

Військова логістика є одним із ключових елементів забезпечення боєздатності армії, оскільки ефективність постачання ресурсів, транспортування обладнання та своєчасного підвезення боєприпасів безпосередньо впливає на результат бойових операцій [1–3]. Сучасні конфлікти характеризуються високою динамічністю, асиметричними загрозами, швидкими змінами тактичної обстановки та необхідністю забезпечення оперативного реагування у режимі реального часу. У таких умовах традиційні методи логістичного планування, такі як статичні маршрути, централізоване управління та жорстко фіксовані графіки, виявляються малоефективними або навіть непридатними. Використання безпілотних літальних апаратів (БпЛА) відкриває нові можливості для вирішення логістичних задач. Завдяки мобільності, здатності працювати в умовах обмеженого доступу та зниженому ризику для особового складу, дрони вже успішно застосовуються для доставки невеликих, але критично важливих вантажів: медикаментів, комплектів зв'язку, деталей для ремонту техніки. Проте інтеграція БпЛА у військову логістику вимагає вирішення ряду складних проблем, пов'язаних із маршрутизацією, управлінням енергетичними ресурсами, уникненням загроз та координацією груп дронів у динамічних і небезпечних умовах [4, 5].

Одним з найбільш перспективних підходів до вирішення цих задач є навчання з підкріпленням (Reinforcement Learning, RL), яке дозволяє агенту (у нашому випадку БпЛА або групі БпЛА) навчатися оптимальної поведінки на основі взаємодії з середовищем, максимізуючи довгострокову винагороду. Особливої уваги заслуговують модифікації RL, такі як багатоагентне підкріплювальне навчання (Multi-Agent RL, MARL), що забезпечує координацію між кількома дронами, а також підходи з обмеженнями (Constrained MDP, CMDP), які дозволяють враховувати обмеження безпеки, енергоспоживання та зон доступу. Проблематика стає ще складнішою у зв'язку з наявністю стохастичних факторів: зміни погодних умов, непередбачуваних загроз, ворожих дій, обмеженої пропускної здатності каналів зв'язку. Це обумовлює необхідність використання ризик-орієнтованих методів (наприклад, CVaR Conditional Value at Risk), що дозволяють мінімізувати негативні наслідки екстримальних подій, критичних для військових операцій [6–8].

Таким чином, наукове завдання полягає у систематизації сучасних методів RL, які можна ефективно застосовувати у військовій логістиці з використанням БпЛА, з урахуванням як математичних моделей, так і алгоритмічних рішень, здатних забезпечити надійну та безпечну роботу у реальних бойових умовах.

Аналіз останніх досліджень і публікацій

За останнє десятиліття інтерес до використання підкріплювального навчання у військовій логістиці та робототехніці значно зріс, що підтверджується зростаючою кількістю публікацій у провідних журналах з інтелектуальних систем, штучного інтелекту та автономних роботів. Розглянемо основні напрями досліджень, які формують наукову основу даної теми.

Перший пласт робіт пов'язаний із марковськими процесами прийняття рішень (MDP) для планування маршрутів автономних апаратів у складних середовищах. Класичні роботи, такі як дослідження [1], заклали базу RL, але вже у 2010-х з'явилися роботи, що демонструють практичну реалізацію цих підходів для літаючих апаратів. Наприклад, у [2] представлено застосування Q-learning та SARSA для навігації БпЛА з урахуванням перешкод та обмеженого запасу енергії. Ця робота показала, що навіть табличні методи RL здатні забезпечувати адаптивне планування в умовах змінної обстановки, хоча масштабування на великі простори станів залишалося проблемою.

Другий напрям досліджень – багатоагентне підкріплювальне навчання (MARL), яке дозволяє координувати групи дронів для спільного виконання завдань. У роботі [3] запропоновано Multi-Agent Deep Deterministic Policy Gradient (MADDPG), що забезпечує стабільне навчання у сценаріях з частковою спостережуваністю. Подальші дослідження, зокрема [4], інтегрували цей підхід із механізмами централізованого навчання та децентралізованого виконання (CTDE), що особливо важливо у військовій логістиці, де під час виконання завдань зв'язок між дронами може бути обмеженим або може переривatися.

Третій вагомий напрям це застосування Constrained Markov Decision Processes (CMDP). У роботі [5] було запропоновано алгоритм Constrained Policy Optimization (CPO), який забезпечує дотримання заданих обмежень (наприклад, рівня ризику або витрати енергії) при оптимізації політики. Подальші дослідження [6] показали ефективність CMDP у сценаріях, де критично важливо уникати загроз, таких як ворожі засоби ППО.

Важливим підходом до підвищення надійності є ризик-орієнтоване підкріплювальне навчання. В роботі [7] автори продемонстрували використання Conditional Value at Risk (CVaR) для мінімізації ймовірності катастрофічних подій. У роботі [8] запропоновано поєднання CVaR із моделями прогнозування загроз, що дало змогу дронам вибирати маршрути, які зменшують імовірність потрапляння в зону ураження навіть у випадках неповної інформації.

Окрему групу досліджень складають роботи з Graph Neural Networks (GNN) для моделювання логістичних мереж. В статті [9] показали, що GNN можуть ефективно узагальнювати політики на нові топології мережі, що критично важливо для військових логістичних операцій, де структура транспортних шляхів змінюється внаслідок бойових дій. Це дозволяє швидко перебудовувати маршрути без повторного тривалого навчання.

Нарешті, практична інтеграція RL у військові симуляційні середовища розглянута у [10], де запропоновано використання підходу sim-to-real з доменними випадковостями для підвищення стійкості політик при перенесенні їх з симулятора у реальний світ. Автори показали, що цей підхід знижує падіння ефективності при переході від навчання у віртуальному середовищі до роботи у фізичних умовах на 30–40 %.

Формулювання мети дослідження

Метою цієї роботи є огляд наукових підходів до застосування методів навчання з підкріпленням у завданнях військової логістики із використанням безпілотних літальних апаратів. У роботі узагальнюються формальні моделі прийняття рішень (MDP, CMDP, Dec-POMDP), розглядаються підходи глибинного та багатоагентного RL, а також методи, орієнтовані на безпеку та управління ризиками.

Основними завданнями дослідження є:

- систематизація існуючих підходів та класифікація їх застосування до різних типів логістичних задач;
- аналіз переваг і обмежень сучасних алгоритмів RL у контексті військових умов;
- формулювання узагальнених висновків щодо придатності методів RL до практичного використання у військовій логістиці.

Об’єктом дослідження виступає процес прийняття рішень у військових логістичних системах з використанням БпЛА, а предметом є алгоритми RL та їх здатність забезпечувати ефективне і безпечне планування дій у динамічному середовищі.

Викладення основного матеріалу дослідження

Проблема інтеграції RL у військову логістику з використанням БпЛА вимагає одночасного врахування як алгоритмічних, так і інженерних аспектів. У центрі цієї задачі є необхідність адаптивного планування та координації в умовах невизначеності, ризику та жорстких обмежень ресурсів. На відміну від цивільних сценаріїв, де оптимізація часто зводиться до мінімізації часу чи витрат, у військових умовах критичними є збереження апаратів, виконання місії та зниження операційних ризиків навіть за рахунок збільшення витрат часу або енергії [3, 5, 7].

Базова модель RL для окремого БпЛА може бути представлена як марковський процес прийняття рішень (MDP), що описується кортежем:

$$M = \langle S, A, P, R, \gamma \rangle, \tag{1}$$

де S – множина можливих станів (позиція дрона, заряд батареї, доступність каналів зв’язку, інформація про загрози); A – множина допустимих дій (зміна курсу, висоти, швидкості, виконання доставки); $P(s' \mid s, a)$ – ймовірність переходу в стан s' після виконання дії a в стані s; $R(s, a)$ – функція винагороди; $\gamma \in (0, 1)$ – коефіцієнт дисконтування, що відображає пріоритет короткострокових або довгострокових вигод. У задачах військової логістики важливо враховувати обмеження, які природно інтегруються у модель Constrained MDP (CMDP):

$$M_c = \langle S, A, P, R, C, d, \gamma \rangle, \tag{2}$$

де $C(s, a)$ – функція витрат (наприклад, споживання енергії або ризик перебування в зоні ураження), а d – допустимий рівень цих витрат. Оптимізація у CMDP здійснюється з урахуванням обмежень:

$$\max_{\pi} \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right], \quad \text{де} \quad \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right] \le d. \tag{3}$$

Методи на кшталт Constrained Policy Optimization (CPO) [5] дозволяють дотримуватися цих обмежень під час навчання політики, що особливо цінно для місій, де перевищення допустимого рівня ризику неприпустиме.

Військові середовища за своєю природою є стохастичними: поведінка противника, погодні умови, стан інфраструктури змінюються непередбачувано. Класичне RL намагається максимізувати середню винагороду, але в умовах бойових операцій це може призвести до небажаних сценаріїв: політика, що є оптимальною може бути фатальною в рідкісних, але критично небезпечних випадках. Тому застосовують ризик-орієнтовані критерії, зокрема Conditional Value at Risk (CVaR) [7,8], який формалізується як:

$$CVaR_{\alpha}(Z) = \mathbb{E}[Z \mid Z \le q_{\alpha}(Z)], \tag{4}$$

де $q_{\alpha}(Z)$ – α-квантиль розподілу виграшу Z. Ця метрика дозволяє оптимізувати політику так, щоб мінімізувати негативні наслідки у найгірших α % сценаріїв, що напряму відповідає вимогам військових операцій.

Координація кількох БпЛА є ще більш складним завданням, оскільки необхідно враховувати як спільну мету, так і індивідуальні обмеження кожного апарата. Multi-Agent RL (MARL) пропонує підходи, що поєднують

централізоване навчання з децентралізованим виконанням (CTDE), як у алгоритмах MADDPG [3] або QMIX [4]. У таких сценаріях функція винагороди може мати вигляд:

$$R_t = \frac{1}{N}\sum_{i=1}^{N} r_t^i - \lambda \cdot Collisions_t \tag{5}$$

де перший доданок описує середній прогрес усіх агентів, а другий – штраф за конфлікти, зіткнення або порушення зони безпеки. Важливим аспектом MARL у військовій логістиці є можливість розподілу завдань: одні дрони можуть виконувати доставку, інші – забезпечувати розвідку чи ретрансляцію сигналу зв'язку [6, 10]. У цьому випадку алгоритм повинен враховувати гетерогенність агентів, що підвищує складність оптимізації.

Маршрутизація в умовах бойових дій часто має мережеву природу: точки постачання, проміжні склади, транспортні коридори утворюють граф, вершинами якого є логістичні вузли, а ребрами – можливі шляхи. Graph Neural Networks (GNN) [9] дозволяють кодувати структуру цього графа і вбудовувати її в процес прийняття рішень RL-агентом. Це дає змогу швидко адаптувати політику до змін мережі без повного перенавчання.

Однією з ключових проблем застосування RL у реальних БпЛА є перенесення політик, навчених у симуляторах, у фізичний світ. Через спрощення моделі середовища у симуляторах політики можуть втрачати ефективність у реальних умовах. Технологія domain randomization [10] пропонує під час навчання в симуляторі варіювати параметри середовища (наприклад, силу вітру, масу вантажу, затримки сенсорів), щоб агент навчився працювати у більш широкому спектрі умов. Це значно підвищує стійкість політики при реальному виконанні.

У військовій логістиці критично важливою є оптимізація енергоспоживання. Функція витрат $C(s, a)$ у CMDP може бути пов'язана з моделлю енергоспоживання дрона:

$$E_{total} = E_{hover} \cdot t_{hover} + E_{cruise} \cdot t_{cruise} + E_{maneuver} \cdot t_{maneuver}, \tag{6}$$

де кожен доданок описує витрати енергії у відповідному режимі польоту. Інтеграція цієї моделі у процес оптимізації дозволяє формувати маршрути, що мінімізують ризик передчасного завершення місії через розряд акумулятора.

Використання БпЛА у військовій логістиці передбачає постійну взаємодію з агресивним середовищем, де присутні цілеспрямовані дії противника, спрямовані на виведення апаратів з ладу. Це може включати фізичне ураження (засоби ППО, снайперські постріли), радіоелектронну боротьбу (глушіння та спуфінг GPS), кібернетичні атаки на системи управління. У термінах RL ці загрози можна моделювати як стохастичні переходи у моделі MDP:

$$P'(s' \mid s, a) = (1 - \eta)P(s' \mid s, a) + \eta Q(s' \mid s, a), \tag{7}$$

де η – інтенсивність втручання противника, а Q – розподіл станів, що виникають внаслідок атак. У цьому випадку класична оптимізація середньої винагороди є недостатньою, оскільки політика повинна бути стійкою до несприятливих умов. З цією метою у літературі [7, 8] активно розвивається напрям robust RL, у якому оптимізація політики формулюється як:

$$\pi^* = \arg\max_{\pi} \min_{P' \in P} \mathbb{E}_{p'\pi}\left[\sum_{i=0}^{\infty} \gamma^t R(s_t, a_t\right], \tag{8}$$

де P – множина можливих збурених моделей переходів. Такий підхід дозволяє зменшити ризик катастрофічних відмов, навіть якщо противник змінює умови середовища.

У багатьох бойових сценаріях критично важливим є збереження комунікаційного каналу між дронами та командним центром. RL-агенти можуть навчатися вибирати такі траєкторії, які не лише оптимізують доставку, але й підтримують зв'язок, використовуючи релейні дрони як мобільні ретранслятори. У цьому випадку модель винагороди може містити додатковий член:

$$R_{comm} = -\mu \cdot \max(0, L_{\max} - d_{link}), \tag{9}$$

де $L_{\max}$ – допустима дальність зв'язку, d_{link} – відстань між вузлами мережі, μ – коефіцієнт штрафу. Це стимулює агентів залишатися в межах зони зв'язку, навіть якщо це збільшує шлях доставки. Дослідження [6, 10] показують, що таке поєднання логістичних і комунікаційних обмежень особливо ефективне у координації гетерогенних флотів БпЛА.

В умовах бойових дій попередньо обчислена траєкторія може стати неактуальною вже через кілька хвилин, наприклад, через знищення мосту, зміну ворожих позицій або поява нових зон ураження. Тут на перший план виходить online-RL навчання та meta-RL. Meta-RL [4, 9] дозволяє агенту швидко адаптувати політику до нових умов, використовуючи попередній досвід. Формально це описується як оптимізація початкових параметрів θ політики π_θ, які можна швидко донавчити для нового завдання T_i:

$$\theta^* = \arg\max_{\theta} \sum_{T_i \sim p(T)} \mathbb{E}_{\pi_{\theta'}}\left[R\right], \quad \partial e\, \theta' = \theta - \alpha \nabla_\theta L_{T_i}\left(\pi_\theta\right). \tag{10}$$

Цей підхід дає змогу перебудовувати логістичну мережу у відповідь на зміну лінії фронту або інфраструктури.

Реальні бойові випробування RL-політик є небезпечними та дорогими, тому використання симуляторів є обов'язковим етапом. Проте симулятори повинні відображати не лише фізику польоту, але й моделювати логістичні процеси та поведінку противника. Деякі роботи [10] інтегрують симулятори, де RL-агенти одночасно вирішують задачі:

- вибору маршруту з урахуванням ризиків;
- координації з іншими дронами;
- підтримки зв'язку;
- економії енергії.

Після етапу симуляцій проводиться обмежене тестування в польових умовах, що зменшує розрив між симуляцією та реальністю.

Висновки

Проведений огляд наукових досліджень показав, що навчання з підкріпленням є перспективним інструментом для вирішення задач військової логістики з використанням безпілотних літальних апаратів. Його ключові переваги полягають у здатності до автономного прийняття рішень у складних та динамічних умовах, адаптації до непередбачуваних змін обстановки, координації дій у багатоагентних системах та оптимізації використання обмежених ресурсів, зокрема енергетичних. У роботі систематизовано підходи до моделювання логістичних задач у термінах MDP та CMDP, розглянуто методи ризик-орієнтованої оптимізації політик, інтеграцію прогнозних моделей загроз, використання багатоагентного навчання та гібридних систем управління. Особливу увагу приділено симуляційним середовищам і підходу sim-to-real, який дозволяє мінімізувати розрив між віртуальним навчанням і реальною експлуатацією.

Перспективними напрямами подальших досліджень є:

- розробка більш точних симуляційних моделей, що враховують як фізичні, так і тактичні фактори;
- інтеграція RL із прогнозними та аналітичними модулями для підвищення ефективності;
- створення гібридних архітектур, що поєднують гнучкість RL із надійністю класичних методів управління;
- застосування методів meta-RL для швидкої адаптації до нових умов операцій.

Список використаної літератури

1. Sutton R. S., Barto A. G. Reinforcement Learning: An Introduction. Cambridge : MIT Press, 2018. 548 p.
2. Abbeel P., Coates A., Ng A. Y. Autonomous helicopter aerobatics through apprenticeship learning. International Journal of Robotics Research. 2010. Vol. 29, No. 13. P. 1608–1639. DOI: 10.1177/0278364910371999
3. Lowe R., Wu Y., Tamar A., Harb J., Abbeel P., Mordatch I. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. Advances in Neural Information Processing Systems (NeurIPS). 2017. P. 6379–6390.
4. Foerster J., Farquhar G., Afouras T., Nardelli N., Whiteson S. Counterfactual multi-agent policy gradients. Proceedings of the AAAI Conference on Artificial Intelligence. 2018. Vol. 32, No. 1. P. 2974–2982.
5. Achiam J., Held D., Tamar A., Abbeel P. Constrained Policy Optimization. Proceedings of the 34th International Conference on Machine Learning (ICML). 2017. P. 22–31.
6. Cohen M. H., Belta C. Safe exploration in model-based reinforcement learning using control barrier functions. Automatica. 2023. Vol. 147. Art. 110684. DOI: 10.1016/j.automatica.2022.110684
7. Chow Y., Tamar A., Mannor S., Pavone M. Risk-sensitive and robust decision-making: a CVaR optimization approach. Advances in Neural Information Processing Systems (NeurIPS). 2015. P. 1522–1530.
8. Chen S., Mo Y., Wu X., Xiao J., Liu Q. Reinforcement Learning-Based Energy-Saving Path Planning for UAVs in Turbulent Wind. Electronics. 2024. Vol. 13, No. 16. Art. 3190. DOI: 10.3390/electronics13163190
9. Khalil E., Dai H., Zhang Y., Dilkina B., Song L. Learning combinatorial optimization algorithms over graphs. Advances in Neural Information Processing Systems (NeurIPS). 2017. P. 6348–6358.
10. Akkaya I., Andrychowicz M., Chociej M., et al. Solving Rubik's Cube with a Robot Hand. arXiv preprint. arXiv:1910.07113. 2019.

References

1. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.
2. Abbeel, P., Coates, A., & Ng, A. Y. (2010). Autonomous helicopter aerobatics through apprenticeship learning. *International Journal of Robotics Research, 29*(13), 1608–1639. https://doi.org/10.1177/0278364910371999
3. Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 6379–6390).
4. Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., & Whiteson, S. (2018). Counterfactual multi-agent policy gradients. *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1, pp. 2974–2982).

5. Achiam, J., Held, D., Tamar, A., & Abbeel, P. (2017). Constrained Policy Optimization. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 22–31).

6. Cohen, M. H., & Belta, C. (2023). Safe exploration in model-based reinforcement learning using control barrier functions. Automatica, 147, 110684. https://doi.org/10.1016/j.automatica.2022.110684

7. Chow, Y., Tamar, A., Mannor, S., & Pavone, M. (2015). Risk-sensitive and robust decision-making: a CVaR optimization approach. In *Advances in Neural Information Processing Systems* (Vol. 28, pp. 1522–1530).

8. Chen, S., Mo, Y., Wu, X., Xiao, J., & Liu, Q. (2024). Reinforcement Learning-Based Energy-Saving Path Planning for UAVs in Turbulent Wind. Electronics, 13(16), 3190. https://doi.org/10.3390/electronics13163190

9. Khalil, E., Dai, H., Zhang, Y., Dilkina, B., & Song, L. (2017). Learning combinatorial optimization algorithms over graphs. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 6348–6358).

10. Akkaya, I., Andrychowicz, M., Chociej, M., et al. (2019). Solving Rubik's Cube with a Robot Hand. *arXiv*. https://arxiv.org/abs/1910.07113

Дата першого надходження рукопису до видання: 25.09.2025
Дата прийнятого до друку рукопису після рецензування: 22.10.2025
Дата публікації: 28.11.2025