УДК 004.89 **doi: 10.32620/reks.2026.2.12**

М. О. КУШНАРЬОВ

Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна

МЕТОД БЕЗПЕЧНОГО ГЛИБОКОГО НАВЧАННЯ З ПІДКРІПЛЕННЯМ ДЛЯ ГАРАНТОВАНОГО ДОТРИМАННЯ ФІЗИЧНИХ ОБМЕЖЕНЬ В АВТОНОМНИХ ЕНЕРГОСИСТЕМАХ КРИТИЧНОЇ ІНФРАСТРУКТУРИ (НА ПРИКЛАДІ МЕДИЧНИХ ЗАКЛАДІВ)

Предметом *дослідження є онтологічно-орієнтовані підходи до видобування бізнес-інформації з неструктурованих вебджерел.* ***Метою*** *статті є аналіз ефективності сучасних онтологічно-орієнтованих підходів до видобування бізнес-інформації з неструктурованих вебджерел та обґрунтування їх доцільності для використання в системах підтримки прийняття рішень.* ***Завдання:*** *проаналізувати основні проблеми збору, видобування та оброблення бізнес-інформації з неструктурованих вебджерел; визначити можливості використання онтологій у задачах видобування та інтеграції бізнес-інформації; виконати порівняльний аналіз сучасних онтологічно-орієнтованих підходів та визначити перспективні напрями їх подальшого використання. У ході дослідження застосовано* ***методи*** *аналізу та узагальнення наукових джерел, системного аналізу, порівняльного аналізу, а також підходи семантичного та онтологічного моделювання. У* ***результаті*** *встановлено, що основними чинниками, які ускладнюють видобування бізнес-інформації з вебпростору, є гетерогенність форматів і структур даних, неоднозначність природної мови, динамічність інформаційного середовища, неповнота та суперечливість відомостей. Обґрунтовано, що використання онтологій забезпечує семантизацію, структуризацію, логічне узгодження й інтеграцію бізнес-інформації в межах корпоративної бази знань, а також створює підґрунтя для підвищення якості аналітичного опрацювання даних. Систематизовано сучасні онтологічно-орієнтовані підходи, зокрема методи на основі шаблонів, глибокого лінгвістичного аналізу та машинного навчання. За результатами порівняльного аналізу встановлено, що найбільш перспективними для систем підтримки прийняття рішень є гібридні підходи, які поєднують переваги різних груп методів, забезпечуючи вищу повноту, гнучкість та семантичну узгодженість результатів.* ***Висновки****. Наукова новизна отриманих результатів полягає в узагальненні та порівнянні сучасних онтологічно-орієнтованих підходів до видобування бізнес-інформації з неструктурованих вебджерел з урахуванням їх придатності до семантичного узгодження даних, формування корпоративної бази знань та використання в системах підтримки прийняття рішень.*

Ключові слова: *енергетична стійкість; Safe DRL; функції Ляпунова; обмежені марковські процеси; Edge-Fog ap-хітектура; мінімізація ризиків; глибоке навчання; критична інфраструктура; мікромережі*

1. Вступ

Сучасні автономні енергетичні системи та кіберфізичні комплекси критичної інфраструктури переживають фазу інтенсивної цифровізації, що супроводжується інтеграцією Інтернету речей (IoT) та інтелектуальних алгоритмів управління на базі штучного інтелекту (AI) [1]. Такі системи широко застосовуються у різних галузях, наприклад, в енергетиці, транспортних комплексах та об'єктах охорони здоров'я, зокрема, на виробничих підприємствах авіабудівного профілю.

Функціонування цих високотехнологічних систем критично залежить від стабільного енергопостачання, особливо в умовах аварійних ситуацій, техногенних катастроф або навмисного пошкодження енергетичної інфраструктури. У таких умовах виникає необхідність розробки нових підходів до інтелектуального управління енергоресурсами, орієнтованих на забезпечення фізичної стійкості (resilience) системи, а не лише на оптимізацію енергоспоживання.

Запропонований підхід також є релевантним для систем енергоменеджменту в авіаційно-космічній техніці, зокрема для бортових енергосистем літальних апаратів, безпілотних комплексів та орбітальних платформ, де критичною є вимога безперервності живлення та дотримання фізичних обмежень у реальному часі.

1.1. Мотивація дослідження

Якщо в мирний час фокус досліджень лежав переважно на енергоефективності та оптимізації

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

витрат [2, 3], то в умовах воєнного стану або техногенних катастроф ключовим пріоритетом стає забезпечення фізичної живучості (resilience) автономних енергетичних систем та безперервності функціонування об'єктів критичної інфраструктури. Зростання глобальної нестабільності посилює увагу до розробки стійких мікромереж (Resilient Microgrids), особливо для об'єктів із підвищеними вимогами до надійності, зокрема в енергетиці, транспортних системах та медичних закладах. Дослідження підтверджують, що системи управління енергоресурсами в таких умовах мають бути оптимізовані не для економічного прибутку, а для максимізації «Днів Автономії» (Days of Autonomy) та мінімізації впливу відключення на функціональну спроможність об'єкта [4].

1.2. Огляд літератури

Традиційні системи управління енергоспоживанням у будівлях (BEMS/HEMS) орієнтовані переважно на роботу в умовах стабільної експлуатації. Вони використовують статичні алгоритми перемикання або прості прогнозні моделі, спрямовані на балансування навантаження та досягнення економічної ефективності. В умовах аварійних або екстремальних режимів, таких як раптовий повний блекаут або робота від генераторів з обмеженим ресурсом палива, такі системи демонструють критичну нездатність до адаптивного та прогностичного управління. Особливою вразливістю таких систем є залежність від централізованих обчислювальних інфраструктур, включаючи хмарні сервіси. Втрата зв'язку з глобальною мережею (Single Point of Failure, SPOF) [5] під час блекауту призводить до повної нефункціональності інтелектуальних компонентів системи, що є неприпустимим для об'єктів критичної інфраструктури.

Ряд наукових робіт присвячено застосуванню методів глибокого навчання з підкріпленням (DRL) для оптимізації енергоспоживання [1]. Однак більшість існуючих підходів орієнтовані на побутовий сектор, комерційні будівлі або загальні мікромережі з акцентом на зниження витрат. Проблема полягає в тому, що тренувальний процес на основі методів навчання з підкріпленням без вбудованих обмежень безпеки (model-free RL) значною мірою покладається на випадкові рішення для дослідження середовища (exploration), що може призвести до генерації дій, які порушують критичні обмеження безпеки та спричиняють катастрофічні наслідки [6].Через відсутність теоретичних гарантій безпеки в енергосистемах, пряме впровадження класичних алгоритмів RL у реальні автономні енергетичні системи вважається обмеженим або неприйнятним.

1.3. Мета та завдання дослідження

Розвиваючи результати попередніх досліджень [2], у яких було обґрунтовано ефективність стандартних DRL-архітектур для задач адаптивного управління, у даній роботі здійснюється перехід від класичного навчання з підкріпленням до парадигми Safe Reinforcement Learning із жорсткими обмеженнями безпеки [6]. Якщо раніше фокус досліджень був спрямований переважно на максимізацію очікуваної винагороди, то в даній роботі пріоритетом є гарантоване дотримання фізичних обмежень системи у будь-який момент часу через інтеграцію методів теорії автоматичного керування [7].

Метою роботи є розробка методу безпечного глибокого навчання з підкріпленням для задач інтелектуального управління автономними енергетичними системами критичної інфраструктури в умовах невизначеності та обмежених ресурсів.

Наукова новизна полягає у розробці гібридного підходу, що поєднує методи глибокого навчання з підкріпленням із інструментами теорії керування для забезпечення формальних гарантій безпеки під час прийняття рішень у реальному часі.

Для досягнення поставленої мети у роботі вирішуються такі завдання:

1. Формалізація задачі управління енергосистемою як обмеженого марковського процесу (CMDP) для явного врахування фізичних обмежень.
2. Розробка математичної моделі з інтеграцією безпечного шару (Safety Layer) на основі функцій Ляпунова для гарантування дотримання обмежень у процесі навчання та експлуатації.
3. Побудова децентралізованої Edge-Fog архітектури для підвищення живучості системи та зменшення залежності від хмарних сервісів.
4. Експериментальна перевірка ефективності запропонованого підходу у спеціалізованому середовищі HospitalEnergyEnv для сценаріїв тривалого блекауту.

2. Аналіз стану проблеми та обґрунтування наукової новизни дослідження

Сучасний етап розвитку інтелектуальних систем управління енергією (EMS) характеризується зміною пріоритетів від чистої економічної вигоди до забезпечення фізичної стійкості (resilience) критичних об'єктів.

Традиційні model-free методи навчання з підкріпленням, такі як DQN або PPO, покладаються на випадкове дослідження середовища (exploration) [6]. Аналіз стану проблеми показує, що такий підхід призводить до прийняття «поганих» рішень під час навчання, які в реальних енергосистемах можуть спричинити каскадні відмови або незворотне пошкодження дороговартісного обладнання (BESS, генераторів). Більшість існуючих робіт у галузі «розумних будинків» ігнорують цю проблему, припускаючи наявність безпечного симулятора. Проте для автономних кіберфізичних систем критичної інфраструктури, де система повинна адаптуватися до унікальних умов функціонування в реальному часі (online learning), відсутність теоретичних гарантій безпеки є головним бар'єром для впровадження засобів AI [8, 9].

У нашій попередній роботі обмеження системи інтегрувалися у функцію винагороди як вагові коефіцієнти (w_4 за порушення SoC_{min}) [2]. Цей підхід (Lagrangian relaxation) є поширеним, проте аналіз досліджень 2024-2025 років свідчить про його недостатність: методи Лагранжа забезпечують виконання обмежень лише асимптотично (після повної збіжності моделі) і не дають жодних гарантій безпеки в процесі тренування [6].

Для вирішення цієї проблеми у даному дослідженні здійснено перехід до парадигми обмежених марковських процесів прийняття рішень (CMDP). На відміну від стандартного MDP, у CMDP явно відокремлено цільову функцію максимізації винагороди від функції вартості безпеки (C(s, a)). Це дозволяє застосувати методи проектування дій (a-projection), які математично гарантують, що очікувані витрати на безпеку не перевищать заданий поріг α.

Найперспективнішим напрямком останнього часу стало використання функцій Ляпунова для перетворення глобальних траєкторних обмежень («не допустити колапсу протягом 48 годин») у локальні покрокові умови безпеки. Світові лідери в області Safe RL довели, що побудова Lyapunov-induced safe set дозволяє визначити множину політик, які є апріорі безпечними [7].

Однак, аналіз літератури виявив наукову лакуну: більшість методів на основі Ляпунова орієнтовані на прості роботизовані задачі (наприклад, MuJoCo) і не враховують специфіку автономних енергетичних систем критичної інфраструктури. У даній статті ця лакуна заповнюється через розробку Safety Assessment Optimization Model (SAOM), яка інтегрує фізичні знання про мікромережу та обмеження стандарту NFPA 99 безпосередньо у математичний апарат функцій Ляпунова [3].

Наукова новизна запропонованого методу, порівняно з існуючими рішеннями та нашою попередньою моделлю [2], полягає у:

1. **Методологічній новизні**: Вперше формалізовано задачу живучості медичного закладу як CMDP з інтегрованим шаром a-проекції, що базується на нейронних функціях Ляпунова [7]. Це дозволяє отримати нульову кількість порушень фізичних обмежень вже з першого епізоду навчання, що є критичним для систем із підвищеними вимогами до безпеки.
2. **Технічній новизні:** Удосконалено Edge-Fog архітектуру шляхом впровадження «фільтра безпеки» на базі квадратичного програмування (QP) на рівні Fog-вузла [10]. Це забезпечує мінімально-інвазивне коригування дій агента: система обирає безпечну дію, яка є максимально близькою до оптимальної дії нейромережі в сенсі евклідової метрики.
3. **Концептуальній новизні:** Синхронізація математичної умови дрейфу Ляпунова з вимогами NFPA 110 щодо надійності систем аварійного живлення, що перетворює абстрактну «стабільність алгоритму» на практичну «фізичну живучість об'єкта».

3. Децентралізована Edge-Fog архітектура системи управління енергоресурсами

Для гарантування енергетичної стійкості об'єктів критичної інфраструктури та забезпечення повної функціональної автономності в умовах знеструмлення зовнішньої мережі, у роботі реалізовано трирівневу ієрархічну архітектуру, що базується на парадигмі Edge-Fog Computing [5]. Пропонована структура (рис. 1) усуває критичну вразливість, притаманну традиційним Building Energy Management Systems (BEMS), які зазвичай побудовані на централізованих хмарних обчисленнях і втрачають працездатність при зникненні каналів зв'язку – ситуації, що є типовою під час масштабних блекаутів.

Базовий рівень сприйняття (Perception Layer) складається з гетерогенної мережі інтелектуальних сенсорів та виконавчих пристроїв, інтегрованих безпосередньо з обладнанням та системами об'єкта відповідно до вимог стандарту NFPA 99 [3,11]. Для забезпечення територіального покриття та живучості мережі використано динамічну Mesh-топологію. Вона дозволяє системі автоматично переконфігурувати маршрути передачі даних у разі виходу з ладу окремих вузлів зв'язку. Особлива увага приділена комунікаційним протоколам: для

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

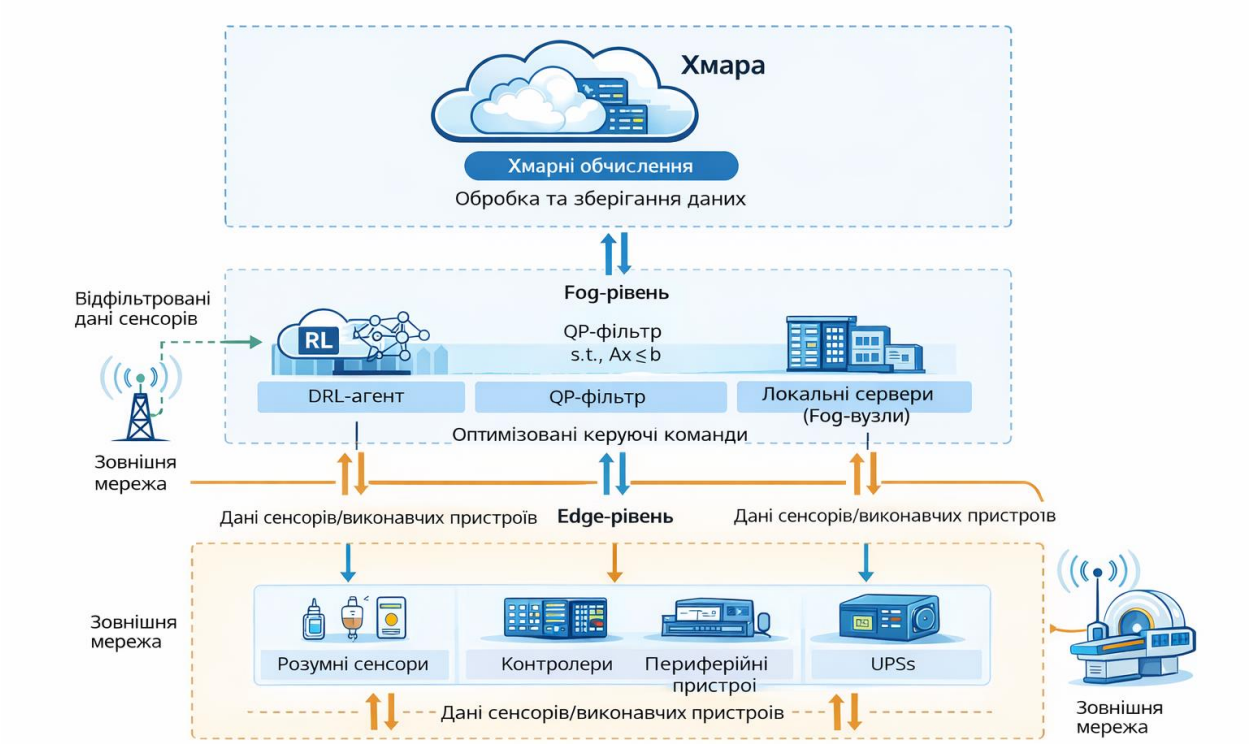

Рис. 1. Децентралізована Edge-Fog архітектура системи управління енергоресурсами шпиталю (зокрема шпиталю)

критичних вузлів обрано стандарт CoAP (Constrained Application Protocol) [12, 13], який завдяки двійковому формату повідомлень та роботі поверх UDP демонструє вищу енергоефективність та менші часові затримки при завершенні транзакцій порівняно з традиційним MQTT. Це критично для підтримки цілісності даних в умовах деградованих мереж із низькою пропускною здатністю [12, 13].

Рівень периферійних обчислень (Edge Intelligence) базується на конвергенції технологій TinyML та вбудованих систем на основі мікроконтролерів класу ARM Cortex-M з об'ємом оперативної пам'яті SRAM від 32 до 512 кБ [5]. Використання апаратних платформ з енергоспоживанням у режимі глибокого сну та інференсу менше 1 мВт дозволяє сенсорним вузлам зберігати повну функціональність від мініатюрних автономних джерел живлення протягом усього періоду блекауту. Основна науково-технічна функція TinyML-моделей на цьому рівні полягає у миттєвій детекції аномалій енергосистеми, таких як раптові флуктуації вольтажу або локальні перевантаження окремих ліній живлення, з часом відгуку менше 10 мс. Окрім швидкодії, локальна обробка забезпечує принциповий рівень приватності даних (Privacy-by-Design): оскільки машинне навчання відбувається безпосередньо біля джерела детальні профілі споживання енергоресурсів та специфічні режими роботи обладнання не передаються до Fog-вузла, що гарантує сувору відповідність міжнародним етичним та юридичним нормам, таким як HIPAA та GDPR [12].

Стратегічна координація та довгострокова оптимізація ресурсів зосереджені на рівні Fog Node, який функціонує як локальний сервер оркестрації в межах захищеного периметра об’єкта. На цьому рівні розгорнуто стратегічного Safe DRL агента, завданням якого є агрегація векторів стану від усіх периферійних вузлів та формування команд для систем аварійного живлення (EPSS). Для підвищення точності стратегічного планування Fog-рівень інтегрує прогностичні модулі на базі LSTM-мереж, які оцінюють динаміку сонячної генерації та очікуваного попиту, що дозволяє агенту діяти антиципово [1]. Важливою архітектурною особливістю є підтримка механізмів федеративного навчання (Federated Learning) [5], що дозволяє розподіленим об’єктам колективно покращувати алгоритми детекції кібератак через обмін вагами моделей без розкриття локальних даних. Для протидії специфічним загрозам, таким як атаки типу False Data Injection (FDI), в архітектуру вбудовано вторинний контур управління, де умови стабільності Ляпунова забезпечують стійку збіжність параметрів напруги та частоти мікромережі навіть у разі маніпуляції зловмисником частиною вхідних

сигналів [4]. Таким чином, запропонована архітектура формує багаторівневий ешелонований бар'єр, де миттєва реакція Edge-рівня гармонійно поєднується зі стратегічною глибиною та кіберстійкістю Fog-серверів, створюючи надійний фундамент для сертифікації автономних систем.

Подібна ієрархічна архітектура може бути застосована в бортових системах керування літальних апаратів, де Edge-рівень відповідає за локальні контролери, а Fog-рівень – за інтегровані обчислювальні модулі управління польотом та енергоресурсами.

Таким чином, дана архітектура забезпечує не лише відмовостійкість, але й створює обчислювальну основу для реалізації алгоритмів безпечного навчання з підкріпленням у реальному часі [6].

4. Методологія Safe DRL: формалізація CMDP та шар Ляпуновської проекції

Необхідність використання гібридного підходу зумовлена тим, що стандартне навчання з підкріпленням без додаткових обмежень демонструє нестабільність на етапах дослідження (exploration). Водночас класичні методи керування забезпечують стабільність, проте позбавлені адаптивності до динамічних змін у стохастичному середовищі функціонування об’єкта. Це вимагає об’єднання підходів для створення системи, де інтелектуальний агент забезпечує стратегічну гнучкість, а математичні фільтри – фізичну безпеку.

Фундаментальною науковою задачею є трансформація процесу навчання агента таким чином, щоб забезпечити «безпечне дослідження» (safe exploration) з нульовою кількістю порушень критичних лімітів уже з першого кроку тренування [6].

На відміну від традиційного навчання з підкріпленням, де метою є лише максимізація очікуваної дисконтованої винагороди $J(\pi)$, управління критичною інфраструктурою вимагає введення множини адитивних обмежень.

Система може бути представлена кортежем

$$\mathcal{M} = \langle S, A, P, R, C, \gamma, \alpha \rangle,$$

де S – множина можливих станів системи;

A – множина допустимих керуючих дій;

P(s^'|s,a) – ймовірність переходу системи у стан s^' при виконанні дії a у стані s;

R(s,a) – функція винагороди, що визначає якість управління;

C(s,a) – функція вартості обмежень (порушення безпеки);

γ – коефіцієнт дисконтування;

$\alpha \in \mathbb{R}_{\geq 0}$ – допустимий бюджет безпеки.

Вектор стану системи x(t): визначає повний набір параметрів (енергетичних, контекстуальних або експлуатаційних) у момент часу t.

Вектор дій a(t) являє собою множину керуючих команд (наприклад, перемикання навантаження, зміну потужності генерації), що подаються на виконавчі механізми. Керування системою розглядається як процес послідовного прийняття рішень, де стратегія π відображає стан x(t) у дію a(t) з метою оптимізації цільового функціоналу при суворому дотриманні безпечних меж стану [7].

У даному контексті функція C(s,a) приймає значення 1, якщо стан системи виходить за межі фізичної безпеки (наприклад, $\mathrm{SoC_t} < \mathrm{SoC_{min}}$), та 0 в іншому випадку. Задача оптимізації формулюється як максимізація цільової функції за умови суворого виконання бюджетного обмеження [2]:

$$\max_{\pi} \quad \mathbb{E}[r(\pi)] \quad \text{s.t.} \quad \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t C(s_t, a_t)\right] \leq \alpha$$

Вибір α=0 обумовлений специфікою задачі, що вимагає абсолютної відсутності критичних розрядів батарей, які можуть призвести до втрати працездатності системи.

Ключова складність застосування CMDP полягає в тому, що траєкторні обмеження є глобальними (залежать від усієї послідовності дій), тоді як агент приймає рішення локально. Для подолання цієї розбіжності ми використовуємо теорію стійкості Ляпунова для перетворення глобальних обмежень у покрокові умови дрейфу. Для заданої базової безпечної політики π_B будується функція Ляпунова $L_{\pi_B}(s)$, яка може бути інтерпретована як апроксиматор очікуваного кумулятивного ризику системи. Оператор Беллмана для вартості обмежень $\mathcal{B}^{\pi,\mathcal{C}}[L](s)$ визначено як:

$$\mathcal{B}^{\pi,\mathcal{C}}[L](s) = \sum_{a} \pi(a|s).$$

Умовою того, що політика π є «не менш безпечною», ніж π_B, є дотримання негативного або нульового дрейфу Ляпунова, що математично гарантує збереження системи всередині інваріантної безпечної множини станів.

Для надання теоретичних гарантій стабільності вводиться скалярна функція Ляпунова V(x), що характеризує ступінь відхилення системи від безпечного стану. Умова стабільності за Ляпуновим задається як:

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

$$V'(x) \leq 0.$$

Це означає, що для стабільної роботи системи обране керування a(t) повинно забезпечувати незростання функції Ляпунова з часом. Для корекції дій RL-агента використовується шар безпеки на базі квадратичного програмування (QP)[10]. Задача мінімізації має вигляд:

$$\min_{a} \frac{1}{2} |a - a_{RL}|^2$$

за умови дотримання обмеження стабільності $\dot{V}(x, a) \leq -\alpha V(x)$. Це гарантує, що фізично виконувана команда завжди відповідає умовам безпеки мікромережі.

Найбільш вагомим внеском даної роботи є впровадження аналітичного шару проекції (a-projection), який діє як «запобіжний фільтр» між нейромережею агента та виконавчими механізмами енергосистеми [1]. Кожна дія a_π~π_θ (·|s), яку пропонує Actor-мережа, піддається мінімально-інвазивному коригуванню. Ми формулюємо задачу знаходження проектованої дії $\tilde{a}$ як задачу квадратичного програмування (QP) в евклідовому просторі:

$$\tilde{a} = \arg\min_{a \in A} \quad \frac{1}{2} |a - a_\pi|^2 \backslash \text{ s.t. } \left(a - a_B(s)\right)^T g_L(s) \leq \hat{\epsilon}(s),$$

де $a_B(s)$ - дія, що генерується консервативною базовою політикою (наприклад, негайне відключення гнучких навантажень); $g_L(s)$ - градієнт Ляпуновської Q-функції вартості $\nabla_a Q_L(s, a)|_{a=a_B(s)}$, який вказує напрямок найшвидшого зростання ризику; $\hat{\epsilon}(s)$ - залишковий бюджет безпеки, що динамічно обчислюється як

$$\hat{\epsilon}(s) = (1 - \gamma)\left(d_0 - D_{\pi_B}(s_0)\right),$$

де d_0 – поріг допустимого ризику.

Для забезпечення можливості наскрізного навчання (end-to-end training) було використано диференційовані QP-рішальники, такі як OptNet [14]. Це дозволяє градієнтам помилки винагороди проходити через шар проекції назад до вагів нейромережі Actor, що значно підвищує стабільність навчання та ефективність вибірки (sample efficiency) порівняно з методами, які базуються на простому відкаті (backtracking). Такий підхід гарантує, що фізично виконувана команда $\tilde{a}$ завжди відповідає умовам безпеки мікромережі, роблячи процес тренування ШІ-агента безпечним для реальних енергетичних систем критичної інфраструктури та для забезпечення безпечного управління енергетичними режимами в авіаційних та космічних системах, де порушення обмежень може призвести до критичних відмов.

5. Формалізація етично-орієнтованої функції винагороди на базі стандарту NFPA 99

У кіберфізичних системах критичної інфраструктури функція винагороди R_t перестає бути суто інженерним інструментом і стає математичним втіленням пріоритетів безпеки та надійності. Наукова новизна запропонованої структури винагороди полягає у переході від економічно-центричної парадигми (мінімізація вартості кВт·год) до парадигми фізичної живучості, де ціннісна вага навантаження визначається категоріями ризику міжнародного стандарту NFPA 99 [11]. Згідно з NFPA 99, навантаження класифікується за чотирма рівнями критичності. У моделі ми формалізуємо цю класифікацію через два вектори попиту: D_{crit} (Категорія 1 – критичні навантаження, відмова яких призводить до тяжких наслідків; зокрема системи життєзабезпечення у шпиталі) та D_{flex} (Категорії 2/3 – допоміжні системи). Етично-орієнтована функція винагороди синтезує ці пріоритети наступним чином:

$$R_t = w_1 \cdot I\left(D_{crit}^{sup} \geq D_{crit}^{req}\right) - w_2 \cdot \left(D_{flex}^{req} - D_{flex}^{sup}\right)^2 - \\ - w_3 \cdot C_{switch} - w_4 \cdot P_{violation}$$

Ключовим аспектом є встановлення домінуючого вагового коефіцієнта w_1 (Reward for Critical Continuity), величина якого на кілька порядків перевищує інші штрафні компоненти ($w_1 \gg w_2, w_3, w_4$) [15]. Це математично гарантує лексикографічну перевагу: агент ніколи не обере стратегію, яка підвищує комфорт (w_2) або зберігає ресурс обладнання (w_3), ціною хоча б мінімального ризику для критичних навантажень (зокрема систем життєзабезпечення). Штраф за дефіцит гнучкого навантаження w_2 (Flexible Deficit Penalty) реалізовано у квадратичній формі, що змушує DRL-агента до рівномірного «плавного» обмеження споживання (Load Curtailment) замість різкого відключення цілих блоків, що мінімізує дискомфорт користувачів системи. Експлуатаційний штраф w_3 (Switching Cost) синхронізований з вимогами стандарту NFPA 110 щодо надійності систем аварійного живлення. Він запобігає «брязканню» (chattering) генераторів та надмірним циклам заряду/розряду BESS, що критично для збереження ресурсу систем EPSS Level 1 у тривалих автономних сценаріях. Таким чином, через механізм винагороди

агент не просто «навчається керувати», а інтерналізує нормативні вимоги безпеки як власну внутрішню логіку, що є обов'язковою умовою для практичного впровадження у системах критичної інфраструктури.

Для валідації було використано сценарій "Blackout-48" (48 годин повної автономії) [6] з такими вихідними даними: джерела генерації – дизель-генератор (100 кВт), сонячні панелі; засоби накопичення – BESS (200 кВт·год); об'єкти споживання – критичні та допоміжні навантаження (зокрема медичне обладнання).

Практична ефективність моделі була підтверджена у сценарії критичного розряду акумуляторних батарей у системі з пріоритетними навантаженнями. Коли Safe DRL агент намагався підтримати роботу другорядних систем (D_{flex}), шар Ляпуновської проекції заблокував цю дію, оскільки вона вела до порушення умови стабільності для критичних навантажень (D_{crit}). Це дозволило зберегти 20% резерву батарей для завершення поточної операції за умов прогнозованого дефіциту сонячної генерації

Результати порівняльного моделювання представлені в таблиці 1.

Таблиця 1

Результати порівняльного моделювання алгоритмів управління енергоресурсами

Алгоритм	$T_{(survival)}$ (год)	Порушення SoCmin	Ефектив-ність (%)
Жадібний (Greedy)	32	> 50 (значне порушення)	66.7
Rule-Based (RB)	41	12	85.4
Safe DRL (запропонований метод)	48	0	100

Графіки динаміки навчання показують, що Safe DRL агент, на відміну від базового DQN, не допускає жодного виходу за межі $SoC(t) \geq 0.3$ вже з першого епізоду завдяки α-проекції [6].

6. Безпека моделей та людиноцентричний підхід (XAI - Explainable Artificial Intelligence)

Система вразлива до атак «отруєння даних» (Data Poisoning) та атак змагального типу (Adversarial Attacks). Впровадження Edge AI дозволяє мінімізувати ці ризики через локальну детекцію аномалій та використання федеративного навчання (Federated Learning), де Fog-вузли обмінюються лише вагами моделей, а не даними [4, 5].

Для довіри користувачів впроваджено модулі XAI, які інтерпретують рішення агента [16]. Наприклад, при відключенні некритичного навантаження система генерує пояснення: «Дія викликана необхідністю збереження резерву енергії для забезпечення безперервності критичних процесів за умов прогнозованого дефіциту генерації» [16].

7. Обговорення результатів

Отримані результати демонструють принципову перевагу запропонованого підходу Safe DRL порівняно з традиційними методами управління енергоресурсами. Зокрема, у сценарії повного знеструмлення ("Blackout-48") базові підходи, такі як жадібні алгоритми (Greedy) та Rule-Based системи, не змогли забезпечити дотримання критичних обмежень рівня заряду акумуляторів, що призводило до потенційно небезпечних ситуацій для критичних навантажень [15].

На відміну від них, запропонована модель продемонструвала нульову кількість порушень обмеження $SoC(t) \geq SoC_{min}$, що підтверджує ефективність інтеграції функцій Ляпунова та механізму а-проєкції для забезпечення безпечного навчання в реальному часі [7]. Це є критично важливим для систем критичної інфраструктури, де навіть одноразове порушення обмежень може мати катастрофічні наслідки. Важливо підкреслити, що досягнутий результат не є лише наслідком консервативної стратегії. На відміну від Rule-Based підходів, які обмежені заздалегідь визначеними сценаріями, Safe DRL агент зберігає здатність до адаптації в умовах невизначеності, зокрема при змінній генерації відновлюваних джерел енергії та варіативному профілі навантаження [6]. Це дозволяє системі не лише уникати критичних станів, але й ефективно використовувати доступні ресурси.

Окремої уваги заслуговує роль децентралізованої Edge-Fog архітектури, яка забезпечує відмовостійкість системи в умовах втрати зв'язку з глобальною мережею. Локалізація обчислень та прийняття рішень на рівні Fog-вузлів усуває критичну точку відмови (Single Point of Failure), притаманну централізованим хмарним рішенням, та дозволяє зберігати працездатність системи навіть у деградованих умовах інфраструктури [5].

Разом з тим, результати дослідження виявляють низку обмежень. По-перше, модель оцінювалася у симуляційному середовищі, що не враховує всі аспекти реальної експлуатації, зокрема деградацію акумуляторів, затримки в комунікаційних мережах та людський фактор. По-друге, використання складних

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

моделей Safe RL потребує значних обчислювальних ресурсів на етапі навчання, що може ускладнювати їх впровадження у системах з обмеженим апаратним забезпеченням [2,17].

Таким чином, результати роботи підтверджують доцільність використання Safe DRL у задачах управління критичною інфраструктурою та відкривають нові можливості для створення стійких, автономних і безпечних енергетичних систем.

8. Висновки

Дане дослідження успішно завершує формування цілісного методологічного та архітектурного базису для інтелектуального управління енергетичними ресурсами об'єктів критичної інфраструктури в екстремальних умовах. Запропонований метод безпечного глибокого навчання з підкріпленням (Safe DRL) дозволяє вперше впровадити адаптивні алгоритми AI у критичні системи з наданням строгих математичних гарантій безпеки. Теоретична значущість роботи полягає у формалізації управління автономною мікромережею об'єкта критичної інфраструктури (зокрема шпиталю) у вигляді обмеженого марковського процесу прийняття рішень (CMDP) з використанням нейронних функцій Ляпунова як сертифікатів стабільності. Розроблений механізм a-проекції вирішує фундаментальну суперечність між необхідністю активного дослідження середовища (exploration) та неприпустимістю порушення фізичних обмежень у реальному часі [6]. Це перетворює навчання з підкріпленням з «чорної скриньки» на прозорий, регуляторно-відповідний інструмент.

Отримані результати підтверджують, що синтез методів машинного навчання та класичної теорії автоматичного керування (QP-фільтрація на базі функцій Ляпунова) дозволяє досягти необхідного балансу між адаптивністю та стабільністю системи. Це робить інтелектуальне управління придатним для критичних систем, де фізичні обмеження (наприклад, SoC_{min}) мають пріоритет над будь-якою оптимізацією.

Практична цінність результатів підтверджена 100% виживаністю критичного обладнання у сценарії 48-годинного блекауту при повній відсутності порушень лімітів заряду акумуляторів від першого кроку навчання. Впроваджена трирівнева децентралізована архітектура Edge-Fog не лише усуває критичну точку відмови, пов'язану з хмарними сервісами, а й забезпечує принципово новий рівень приватності даних через використання TinyML безпосередньо на периферійних вузлах. Перспективи подальших досліджень та масштабування запропонованих рішень включають:

1. Масштабування до Multi-Agent RL (MARL): Розширення моделі на великі багаторівневі комплекси об'єктів критичної інфраструктури, де окремі підсистеми координуються автономними агентами через федеративні протоколи [4].
2. Інтеграцію із засобами Explainable AI (XAI): Використання великих мовних моделей (LLM) як інтерфейсу для пояснення рішень Safe DRL агента користувачам на природній мові, що є ключовим для побудови довіри та сертифікації за стандартами ISO/IEC 42001 [16].
3. Кіберфізичну стійкість до інтелектуальних атак: розробка надійних механізмів захисту від атак типу False Data Injection (FDI) та отруєння даних (Data Poisoning) у розподілених середовищах на основі робастних Ляпуновських стратегій [4].

Таким чином, результати проведеного дослідження утворюють надійний науково-технологічний фундамент для розроблення наступного покоління стійких, автономних та регуляторно-відповідних систем критичної інфраструктури, здатних функціонувати в умовах найвищого ступеня невизначеності та загроз. Запропонований підхід має потенціал застосування не лише в енергосистемах наземних об'єктів, але й у бортових енергосистемах авіаційної та космічної техніки, де критичними є вимоги до відмовостійкості, автономності та безпеки.

Конфлікт інтересів

Автор заявляє, що немає конфлікту інтересів щодо цього дослідження, фінансового, особистого, авторського чи іншого, який міг би вплинути на до слідження та його результати, представлені в цій статті.

Фінансування

Дослідження проводилося без фінансової підтримки.

Доступність даних

Рукопис не містить пов'язаних даних.

Використання штучного інтелекту

Автор підтверджує, що він не використовував технології штучного інтелекту під час створення цієї роботи.

Автор прочитав та погодився з опублікованою версією рукопису.

Література

1. On edge-fog-cloud collaboration and reaping its benefits: a heterogeneous multi-tier edge computing

architecture [Text] / N. Fernando, S. Shrestha, S. W. Loke, & K. Lee // Future Internet. – 2025. – Vol. 17, no. 1. – Article no. 22. DOI: 10.3390/fi17010022.

2. Кушнарьов, М. О. Гібридна модель адаптивної пріоритезації енергетичних ресурсів медичного закладу в умовах критичного дефіциту [Текст] / М. О. Кушнарьов, & І. В. Шостак // Відкриті інформаційні та комп'ютерні інтегровані технології. – 2026. – № 107. – P. 241-256 DOI: 10.32620/oikit.2026.107.16.

3. Safe deep reinforcement learning for microgrid energy management in distribution networks with leveraged spatial-temporal perception [Text] / T. Su, T. Wu, J. Zhao, A. Scaglione, &, L Xie // IEEE Transactions on Smart Grid. – 2023. – Vol. 14, no. 2. – P. 154-165. DOI: 10.1109/TSG.2023.3243142.

4. Safe reinforcement learning with Lyapunov-based constraints for control of an unstable reactor [Text] / J. R. T. Neto, B. D. O. Capron, A. R. Secchi, & A. D. R. Chanona // Systems and Control Transactions. – 2025. – Vol. 4. DOI: 10.69997/sct.137298.

5. Ghasem, M. A comprehensive survey of reinforcement learning: from algorithms to practical challenges [Text] / M. Ghasem, A. H. Moosavi, & D. Ebrahimi // arXiv. – 2025. DOI: 10.48550/arXiv.2411.18892.

6. A review of safe reinforcement learning methods for modern power systems [Text] / T. Su, T. Wu, J. Zhao, A. Scaglione., & L. Xie // Proceedings of the IEEE. – 2025. – Vol. 113, no. 3. – P. 213-255. DOI: 10.1109/JPROC.2025.3584656.

7. A review of safe reinforcement learning: methods, theories, and applications [Text] / S. Gu, L. Yang, Y. Du, G. Chen, F. Walter, & J. Wang // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2024. – Vol. 46, no. 12. – P. 11216-11235. DOI: 10.1109/TPAMI.2024.3457538.

8. Terven, J. R. Deep reinforcement learning: a chronological overview and methods [Text] / J. R. Terven // AI. – 2025. – Vol. 6, no. 3. – Article no. 46. DOI: 10.3390/ai6030046.

9. Prudencio, R. F. A survey on offline reinforcement learning: taxonomy, review, and open problems [Text] / R. F. Prudencio, M. R. O. A. Maximo, & E. L. Colombini // IEEE Transactions on Neural Networks and Learning Systems. – 2024. – Vol. 35. – P. 10237-10257. DOI: 10.1109/TNNLS.2023.3250269.

10. End-to-End Safe Reinforcement Learning through Barrier Functions for Safety-Critical Continuous Control Tasks [Text] / R. Cheng, G. Orosz, R. M. Murray, & J. W. Burdick // Proceedings of the AAAI Conference on Artificial Intelligence. – 2019. – Vol. 33. – P. 3387-3395. DOI: 10.1609/aaai.v33i01.33013386.

11. National Fire Protection Association. NFPA 99: Health Care Facilities Code [Електронний ресурс]. – Quincy: NFPA, 2021. – Режим доступу: https://blog.koorsen.com/overview-of-nfpa-99-health-care-facilities-code – 12.01.2026.

12. A comprehensive survey on impact of applying various technologies on the internet of medical things [Text] / S. E. El-deep, A. A. Abohany, K. M. Sallam, & A. A. Abd El-Mageed // Artificial Intelligence Review. – 2025. – Vol. 58. – Article no. 86. DOI: 10.1007/s10462-024-11063-z.

13. Safe deep reinforcement learning for robust frequency and voltage-constrained networked microgrid restoration [Text] / A. Selim, J. Zhao, J. Dong, & J. Lian // IEEE Transactions on Industry Applications. – 2026. – Vol. 62, no. 2. – P. 3635-3647. DOI: 10.1109/TIA.2025.3626472.

14. Cocault, P. Safe deep reinforcement learning control with self-learned neural Lyapunov functions and state constraints [Text] / P. Cocault, S. Bertrand, & H. Piet-Lahanier // Proceedings of the 10th International Conference on Control, Decision and Information Technologies (CoDIT). – Valletta, Malta, 2024. DOI: 10.1109/CoDIT62066.2024.10708548.

15. Rajagopal, D. AI augmented edge and fog computing for Internet of Health Things (IoHT) [Text] / D. Rajagopal, & P. K. T. Subramanian // PeerJ Computer Science. –2025. – Vol. 11. – Article no. e2431. DOI: 10.7717/peerj-cs.2431.

16. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI [Text] / A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, & et al. // Information Fusion. – 2020. – Vol. 58. – P. 82-115. DOI: 10.1016/j.inffus.2019.12.012.

17. Yang, S. Offline reinforcement learning for microgrid voltage regulation [Text] / S. Yang, & Y. Zhu // arXiv. – 2025. DOI: 10.48550/arXiv.2505.09920.

References

1. Fernando, N., Shrestha, S., Loke, S. W., & Lee, K. On edge-fog-cloud collaboration and reaping its benefits: a heterogeneous multi-tier edge computing architecture. *Future Internet*, 2025, vol. 17, no. 1, article no. 22. DOI: 10.3390/fi17010022.

2. Kushnarov, M. O., & Shostak, I. V. Hibrydna model′ adaptyvnoyi priorytezatsiyi enerhetychnykh resursiv medychnoho zakladu v umovakh krytychnoho defitsytu [Hybrid Model of Adaptive Prioritization of Energy Resources in a Medical Facility under Critical Deficit Conditions]. *Vidkryti informatsiyni ta komp'yuterni intehrovani tekhnolohiyi – Open Information and Computer Integrated Technologies*, 2026, no. 107. pp. 241-256 DOI: 10.32620/oikit.2026.107.16. (In Ukrainian).

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

3. Su, T., Wu, T., Zhao, J., Scaglione, A., & Xie, L. Safe deep reinforcement learning for microgrid energy management in distribution networks with leveraged spatial-temporal perception. *IEEE Transactions on Smart Grid*, 2023, vol. 14, no. 2, pp. 154-165. DOI: 10.1109/TSG.2023.3243142.

4. Neto, J. R. T., Capron, B. D. O., Secchi, A. R., & Chanona, A. D. R. Safe reinforcement learning with Lyapunov-based constraints for control of an unstable reactor. *Systems and Control Transactions*, 2025, vol. 4. DOI: 10.69997/sct.137298.

5. Ghasem, M., Moosavi, A. H., & Ebrahimi, D. A comprehensive survey of reinforcement learning: from algorithms to practical challenges. *arXiv*, 2025. DOI: 10.48550/arXiv.2411.18892.

6. Su, T., Wu, T., Zhao, J., Scaglione, A., & Xie, L. A review of safe reinforcement learning methods for modern power systems. *Proceedings of the IEEE*, 2025, vol. 113, no. 3. pp. 213-255. DOI: 10.1109/JPROC.2025.3584656.

7. Gu, S., Yang, L., Du, Y., Chen, G., Walter, F., & Wang, J. A review of safe reinforcement learning: methods, theories, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, vol. 46, no. 12. pp. 11216-11235. DOI: 10.1109/TPAMI.2024.3457538.

8. Terven, J. R. Deep reinforcement learning: a chronological overview and methods. *AI*, 2025, vol. 6, no. 3. Article no. 46. DOI: 10.3390/ai6030046.

9. Prudencio, R. F., Maximo, M. R. O. A., & Colombini, E. L. A survey on offline reinforcement learning: taxonomy, review, and open problems. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, vol. 35. pp. 10237-10257. DOI: 10.1109/TNNLS.2023.3250269.

10. Cheng, R., Orosz, G., Murray, R. M., & Burdick, J. W. End-to-End Safe Reinforcement Learning through Barrier Functions for Safety-Critical Continuous Control Tasks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, pp. 3387-3395. DOI: 10.1609/aaai.v33i01.33013386.

11. *National Fire Protection Association. NFPA 99: Health Care Facilities Code*. Quincy: NFPA, 2021. Available at: https://blog.koorsen.com/overview-of-nfpa-99-health-care-facilities-code (accessed 12.01.2026)

12. El-deep, S. E., Abohany, A. A., Sallam, K. M., & Abd El-Mageed, A. A. A comprehensive survey on impact of applying various technologies on the internet of medical things. *Artificial Intelligence Review*, 2025, vol. 58, article no. 86. DOI: 10.1007/s10462-024-11063-z.

13. Selim, A., Zhao, J., Dong, J., & Lian, J. Safe deep reinforcement learning for robust frequency and voltage-constrained networked microgrid restoration. *IEEE Transactions on Industry Applications*, 2026, vol. 62, no. 2, pp. 3635-3647. DOI: 10.1109/TIA.2025.3626472.

14. Cocault, P., Bertrand, S., & Piet-Lahanier, H. Safe deep reinforcement learning control with self-learned neural Lyapunov functions and state constraints. *Proceedings of the 10th International Conference on Control, Decision and Information Technologies (CoDIT)*, Valletta, Malta, 2024. DOI: 10.1109/CoDIT62066.2024.10708548.

15. Rajagopal, D., & Subramanian, P. K. T. AI augmented edge and fog computing for Internet of Health Things (IoHT). *PeerJ Computer Science*, 2025, vol. 11. article no. e2431. DOI: 10.7717/peerj-cs.2431.

16. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., & et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 2020, vol. 58. pp. 82-115. DOI: 10.1016/j.inffus.2019.12.012.

17. Yang, S., & Zhu, Y. Offline reinforcement learning for microgrid voltage regulation. *arXiv*, 2025. DOI: 10.48550/arXiv.2505.09920.

Received 20.12.2025, Received in revised form 10.02.2026
Accepted date 15.04.2026, Published date 22.04.2026

SAFE DEEP REINFORCEMENT LEARNING METHOD FOR GUARANTEED COMPLIANCE WITH PHYSICAL CONSTRAINTS IN AUTONOMOUS ENERGY SYSTEMS OF CRITICAL INFRASTRUCTURE (A CASE STUDY OF HEALTHCARE FACILITIES)

Maksym Kushnarov

The study examines the complex processes of intelligent management of energy resilience in modern healthcare facilities during critical situations, large-scale failures, and prolonged outages of the external centralized power supply. The aim is to develop a comprehensive mathematical model and a Safe Deep Reinforcement Learning (Safe DRL) method that ensures guaranteed compliance with the strict physical and operational constraints of hospital energy systems, even during the intensive training phase of a neural network agent. The objectives are: to formalize in detail the decision-making procedure in an energy system in detail by transitioning to the paradigm of Constrained Markov Decision Processes (CMDP); to develop an innovative mathematical model featuring the implementation of a specialized safety layer based on Lyapunov functions; and to ensure high resilience and autonomy of the system

through the implementation of a decentralized Edge-Fog data processing architecture. The methods used include: the theory of Constrained Markov Decision Processes (CMDP), deep reinforcement learning methods based on the Actor-Critic architecture, the mathematical apparatus of Lyapunov stability theory for the analytical correction of actions, and methods of simulation modeling for complex dynamic energy systems. The following results were obtained. In the course of the study, a Safe DRL method was proposed and substantiated, which integrates a Lyapunov-based projection directly into the training loop for the immediate correction of the agent's control actions. This makes it possible to ensure strict theoretical guarantees of maintaining the required State of Charge (SoC) of battery systems and to prevent critical violations of energy system parameters beyond established safety limits, which is crucial for patient life-support. The effectiveness of the proposed approach was confirmed by a series of numerical experiments in the specialized environment, HospitalEnergyEnv. Under a full blackout scenario, the agent demonstrated adaptability and high accuracy in resource management without any violation of the established physical limits during the entire process of autonomous operation. Conclusions. The scientific novelty of the obtained results lies in the following: the existing optimization model for building energy management systems (BEMS) has been improved by introducing an analytical safety projection mechanism, which minimizes the risks of emergency equipment shutdown during the adaptation of artificial intelligence algorithms; further development of decentralized control methods for critical infrastructure based on Edge-Fog computing has been achieved, which significantly increases system fault tolerance in the event of a loss of connection with the global network and ensures obtaining quasi-optimal solutions in high-dimensional problems. The practical value of this work lies in the potential to create highly reliable autonomous energy systems for critical infrastructure facilities.

Keywords: Deep Reinforcement Learning; Lyapunov functions; energy resilience; microgrids; smart hospital; physical constraints; Edge-Fog computing.

Кушнарьов Максим Олександрович – асп. каф. інженерії програмного забезпечення, Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна.

Maksym Kushnarov – PhD Student, the Department of Software Engineering, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine,
e-mail: maksimkushnarov@gmail.com, ORCID: 0009-0003-6322-1740.