УДК: 004.032.26:004.83 DOI: **10.33099/2311-7249/2026-55-1-126-134**

МИКОЛАЙЧУК Роман Антонович,
доктор технічних наук, доцент,
Національний університет оборони України, Київ, Україна,
https://orcid.org/0000-0001-5349-4487

ЛИФАР Олена Іванівна,
Національний університет оборони України, Київ, Україна,
https://orcid.org/0009-0004-4894-2184

МЕТОДОЛОГІЧНІ ЗАСАДИ НЕЙРОМЕРЕЖЕВОЇ СКАЛЯРИЗАЦІЇ БАГАТОКРИТЕРІАЛЬНИХ ЗАДАЧ УПРАВЛІННЯ ДИНАМІЧНИМИ СИСТЕМАМИ ВІЙСЬКОВОГО ПРИЗНАЧЕННЯ

Мета *зводиться до розроблення методологічних засад стан-залежної нейромережевої скаляризації (State-Dependent Neural Scalarization (SDNS)) як універсального диференційованого оператора скаляризації багатокритеріальних цільових функцій для ефективного управління динамічними системами військового призначення.*

Методи дослідження. *Під час проведення дослідження застосовано методи системного аналізу, теорії багатокритеріальної оптимізації, теорії Парето, функціонального аналізу та тензорної алгебри, а також спеціальні методи глибокого навчання і багатоцільового навчання з підкріпленням (Multi-Objective Reinforcement Learning (MORL)), доповнені аналізом диференційованості нелінійних функцій агрегації.*

Отримані результати дослідження. *Показано обмеження класичних методів скаляризації багатокритеріальних задач динамічних систем військового призначення, зокрема, встановлено їх непридатність для роботи з неопуклими множинами Парето та умовами високодинамічного середовища. Формалізовано постановку задачі як перетворення векторної цільової функції у скалярну форму, придатну для інтеграції в алгоритми навчання з підкріпленням. Запропоновано математичну архітектуру стан-залежної нейромережевої скаляризації, що базується на тензорній конкатенації вектора локальних критеріїв та вектора фазового стану системи. Розроблено диференційований оператор агрегації, який забезпечує динамічну зміну пріоритетів критеріїв через механізм граничної норми заміщення як функції стану. Подано алгоритмічну процедуру навчання стан-залежної нейромережевої скаляризації із застосуванням багатоцільового навчання з підкріпленням.*

Елементи наукової новизни. *Вперше запропоновано методологію стан-залежної нейромережевої скаляризації багатокритеріальних задач управління динамічними системами військового призначення, що реалізує перетворення векторної цільової функції у скалярну форму як диференційований оператор, параметризований фазовим станом системи.*

Теоретичне та практичне значення викладеного у статті. *Теоретичне значення зводиться до формалізації стан-залежної нейромережевої скаляризації як диференційованого оператора агрегації критеріїв у задачах багатокритеріального моделювання динамічними системами військового призначення. Запропонований метод забезпечує математично коректне перетворення векторної цільової функції у скалярну форму з урахуванням фазового стану об'єкта, що розширює методологію багатокритеріальної оптимізації та створює основу для інтеграції нейромережевих методів у контури адаптивного керування. Практичне значення зумовлює можливості безпосереднього впровадження методу стан-залежної нейромережевої скаляризації у системи управління типу Command and Control, автономні безпілотні платформи, системи протидії радіоелектронному впливу та автономного кіберзахисту. Використання запропонованого підходу дає змогу забезпечити адаптивне перемикання пріоритетів між ефективністю виконання завдання, безпекою та ресурсними обмеженнями в режимі реального часу, підвищити стійкість алгоритмів навчання з підкріпленням до зміни тактичної обстановки та зменшити ризик некоректної поведінки автономних агентів.*

Ключові слова: *штучний інтелект, машинне навчання, штучні нейронні мережі, навчання з підкріпленням, оцінка ефективності, багатокритеріальна оптимізація, нейромережева скаляризація, теорія управління, обробка даних, управління динамічними системами військового управління.*

Вступ

Постановка проблеми. Сучасний етап розвитку озброєння та військової техніки характеризується інтенсивним впровадженням систем штучного інтелекту (далі – ШІ) у бойові платформи, від автономних безпілотних літальних апаратів (далі – БпЛА) до алгоритмів підтримання прийняття рішень у системах управління та контролю (Command and Control (C2)). У реальних умовах бойових дій функціонування таких систем неминуче стикається з фундаментальною проблемою багатокритеріальності. Процес прийняття рішень вимагає одночасного задоволення низки суперечливих цілей: максимізації

ймовірності ураження цілі, мінімізації ризику для власних сил, економії енергоресурсів та дотримання етичних обмежень щодо мінімізації супутніх втрат [1].

З математичного погляду, задача багатокритеріальної оптимізації (далі – БКО) у динамічних процесах зводиться до пошуку вектора керуючих впливів, що забезпечує оптимальність векторної цільової функції, яка формально записується як пошук екстремуму вектор-функції [2; 3]. Оскільки покращення одного локального критерію (наприклад, швидкості виконання маневру) зазвичай призводить до деградації іншого (приміром, збільшення радіолокаційної або теплової помітності), єдиного глобального екстремуму не існує. З огляду на те, що покращення одного критерію, зазвичай, супроводжується погіршенням іншого, розв'язком задачі багатокритеріальної оптимізації є множина недомінованих альтернатив – фронт Парето [2]. Однак, практичне використання повної множини Парето у задачах багатокритеріальної оптимізації не завжди є можливим. Зокрема, в алгоритмах глибокого навчання з підкріпленням (Reinforcement Learning (далі – RL)), що забезпечують тактичне управління в режимі реального часу, необхідним є вибір єдиного управлінського рішення шляхом скаляризації векторної цільової функції [4].

Традиційні методи скаляризації є статичними, тобто передбачають фіксовані вагові коефіцієнти критеріїв і не враховують миттєву динаміку фазового стану системи. Водночас військові системи належать до стан-залежних (state-dependent) динамічних систем, у яких оптимальність рішення визначається не лише структурою цільової функції, а й поточним вектором стану, що описує оперативну обстановку, ресурсні обмеження, рівень загроз та часовий контекст. У таких системах пріоритети критеріїв мають розглядатися як функції стану, де вагові коефіцієнти адаптивно змінюються залежно від ситуації. Ігнорування цієї властивості призводить до втрати тактичної гнучкості, оскільки агент продовжує оптимізувати фіксовану агреговану функцію навіть у разі суттєвої зміни оперативних умов.

В умовах сучасного бою необхідно забезпечити адаптивне перемикання пріоритетів без розривів частинних похідних у контурі управління, що гарантує керованість і стійкість системи. Це обумовлює доцільність переходу від статичної до стан-залежної нейромережевої скаляризації критеріїв, у якій ваги формуються окремим модулем оцінки ситуації або вбудованим механізмом динамічного перерозподілу вагових коефіцієнтів. Такий підхід дає змогу узгодити багатокритеріальну оптимізацію з природою нелінійних динамічних систем та забезпечити безперервність управлінських рішень у реальному часі.

Аналіз останніх досліджень і публікацій. Питання розроблення методів багатокритеріального управління та оптимізації динамічних процесів є предметом численних наукових досліджень, що демонструють значне методичне різноманіття. Нормативно-правовою основою розвитку та впровадження технологій штучного інтелекту в Україні є Концепція розвитку штучного інтелекту в Україні, а також Постанова Кабінету Міністрів України «Про затвердження плану заходів з реалізації Концепції розвитку штучного інтелекту в Україні на 2021–2024 роки», які визначають основні напрями державної політики, включно з питаннями етичності, безпеки та прозорості алгоритмів. [5]. Дотримання міжнародних стандартів та Політичної декларації про відповідальне використання ШІ у військовій сфері формує засади для забезпечення правової легітимності автономних систем [6].

Значний внесок у розвиток теорії багатокритеріального синтезу та управління зробили вітчизняні вчені. Зокрема, В. Калачов, С. Ткачук, Є. Мерентi та Д. Третяк досліджували проблеми багатокритеріального синтезу ієрархічних структур управління [7]. Питаннями автоматизації процесів та оцінювання якості інформаційного забезпечення в автоматизованих системах займалися О. Крайнов, Р. Грозовський та А. Кравчук [8]. Окремі аспекти динамічного управління та застосування ройового інтелекту для децентралізованих систем висвітлені у працях А. Кучук та В. Терзіяна [9].

Фундаментальні засади нелінійної скаляризації та моделювання множини Парето закладені у працях зарубіжних фахівців. Зокрема, С. Лін та Ц. Чжан запропонували методи гладкої згортки Чебишова (Smooth Tchebycheff Scalarization (далі – STCH)) для градієнтної оптимізації глибоких нейронних мереж [10]. Питання багатоцільового навчання з підкріпленням (Multi-Objective Reinforcement Learning (далі – MORL)) з нелінійними преференціями розкриті Н. Пен та М. Тянь, які обґрунтували методи апроксимації очікуваного скаляризованого повернення [11]. Проблему експлуатації функції винагороди агентом (reward hacking) та методи її нівелювання через алгоритми групи MO-GRPO (Multi-Objective Group Relative Policy Optimization (MO-GRPO)) досліджували Ю. Ічіхарата та Ю. Джіннай [12]. Разом із тим, ці підходи здебільшого пропонують набір статичних вагових схем, залишаючи процес миттєвої адаптації пріоритетів до фазового стану динамічної системи на розсуд розробника.

Слід зазначити, що попередні дослідження доводять: класична лінійна згортка не здатна ефективно апроксимувати неопуклі фронти Парето в умовах високодинамічного бойового середовища. Отже, незважаючи на наявність значного теоретичного доробку, наукове завдання розроблення цілісних, формалізованих методологічних засад стан-залежної нейромережевої згортки (State-Dependent Neural Scalarization (далі – SDNS)), які б інтегрували динаміку фазового стану безпосередньо в оператор агрегації цілей, залишається невирішеним й актуальним.

Мета статті зводиться до розроблення методологічних засад стан-залежної нейромережевої скаляризації (SDNS) як універсального диференційованого оператора скаляризації багатокритеріальних цільових функцій для ефективного управління динамічними системами військового призначення.

Виклад основного матеріалу дослідження

Традиційна математична парадигма перетворення векторної цільової функції у скалярну базується на використанні алгебраїчних агрегаційних функцій, кожна з яких має специфічні обмеження під час застосування у складних динамічних системах. Історично домінуючим підходом є лінійна згортка критеріїв (метод зваженої суми), де глобальна функція ефективності визначається за виразом:

$$J = \sum_{i=1}^{k} w_i f_i(x), \quad (1)$$

де $f_i(x)$ – i-й критерій ефективності;

$w_i \geq 0$ – ваговий коефіцієнт, що визначає відносну важливість i-го критерію ефективності;

k – кількість критеріїв.

Кожному показнику (швидкість, безпека, скритність) призначається статичний ваговий коефіцієнт w_i.

Першочерговим математичним недоліком цього підходу є геометричне обмеження: метод зваженої суми здатний генерувати всі Парето-оптимальні рішення лише за умови опуклості множини досяжних векторів критеріїв. У складних тактичних сценаріях фронт Парето, як правило, має неопуклу структуру. У такому випадку підтримуюча гіперплощина лінійної скаляризації не має точок дотику з внутрішніми неопуклими ділянками, внаслідок чого відповідні Парето-оптимальні рішення залишаються недосяжними [2].

Другим суттєвим обмеженням є постійність граничної норми заміщення MRS_{ij}, що характеризує відносну чутливість функції ефективності до змін критеріїв f_i та f_j, визначається як:

$$MRS_{ij} = \frac{\partial J/\partial f_j}{\partial J/\partial f_i}, \quad (2)$$

де i, j – індекси критеріїв ефективності, що характеризують різні показники системи.

Для лінійної моделі виконується $\frac{\partial J}{\partial f_i} = w_i$:

$$MRS_{ij} = \frac{w_j}{w_i}, \quad (3)$$

тобто відношення граничних ваг є глобальною константою та не залежить від фазового стану системи.

У реальних умовах сучасного бою цінність ресурсу нелінійно та контекстно залежить від тактичної обстановки. Наприклад, за відсутності ворожої протиповітряної оборони пріоритет швидкості виконання місії є високим, тоді як у разі виявлення опромінення радаром супротивника, значущість показників, що впливають на помітність, має зростати непропорційно. Лінійна парадигма принципово не здатна відобразити таку стан-залежну поведінку без запровадження зовнішніх дискретних правил, що призводить до розривів у функції винагороди та її похідних і, відповідно, погіршує аналітичну гладкість контуру управління.

Розвиток нелінійних методів скаляризації, зокрема, мінімаксної згортки Чебишова, що базується на $\mathcal{L}_\infty$-нормі, дає змогу долати обмеження лінійної моделі та знаходити Парето-оптимальні рішення навіть за неопуклої структури фронту. Ізоповерхні такої скаляризації мають гіперкубічну форму, що забезпечує проникнення у внутрішні неопуклі ділянки простору критеріїв [2]. Класична форма Чебишовської скаляризації визначається за виразом:

$$J(x) = \max_i \{w_i | f_i(x) - z_i^* |\}, \quad (4)$$

де z_i^* – референтна (ідеальна) точка, що визначає бажане або найкраще досяжне значення i-го критерію в просторі критеріїв.

Наявність оператора максимуму робить функцію неперервною, але недиференційованою у точках перемикання активного критерію, що ускладнює застосування градієнтних методів оптимізації в нейромережевих системах керування. Хоча існують гладкі апроксимації оператора максимуму (зокрема, STCH), вони зберігають чітко визначену аналітичну форму та не забезпечують адаптивної зміни механізму агрегації відповідно до фазового стану системи.

Усвідомлення фундаментальних обмежень аналітичних функцій скаляризації зумовило використання штучних нейронних мереж як універсальних апроксимаційних моделей багатовимірних конфліктів. Відповідно до теореми універсальної апроксимації Цибенка та Горніка, багатошарові нейронні мережі здатні з довільною точністю апроксимувати неперервні функції на компактних множинах [13; 14], що дає змогу розглядати їх як інструмент моделювання нелінійної структури множини Парето (Pareto Set Modeling).

Подальший розвиток багатоцільового навчання з підкріпленням MORL сприяв формуванню методів прийняття рішень у середовищах із векторною функцією винагороди. У військовій сфері MORL розглядається як перспективний підхід для задач тактичного управління та розподілу ресурсів. Зокрема, у ієрархічних архітектурах управління, RL-агент може приймати високорівневі тактичні рішення, балансуючи між безпекою, оперативною ефективністю та ресурсними обмеженнями.

Крім того, методи MORL (зокрема, проксимальна оптимізація політики (Proximal Policy Optimization, (PPO)), багатокритеріальна оптимізація політики методом PPO (Multi-Objective PPO)) інтегруються в системи автономного кіберзахисту (Autonomous Cyber Defence (ACD)), де агенти функціонують у режимі реального часу, розв'язуючи конфлікт між ізоляцією скомпрометованих вузлів мережі та забезпеченням безперервності роботи критично важливих сервісів управління. Незважаючи на досягнутий прогрес, існуючі підходи зберігають низку нерозв'язаних методологічних проблем:

1. Ігнорування інтеграції динаміки фазового стану, тобто більшість алгоритмів багатоцільового навчання з підкріпленням розглядають вектор критеріїв $F(x)$ та вектор фазового стану s як структурно незалежні величини. У результаті функція скаляризації не містить внутрішнього механізму адаптивної зміни пріоритетів у відповідь на різкі збурення середовища (наприклад, раптове застосування засобів радіоелектронної боротьби). Це призводить до інерційності прийняття рішень та затримки у переорієнтації політики агента.

2. Проблема забезпечення повної диференційованості зводиться до того, що динамічні платформи (ракети, винищувачі, безпілотні системи) функціонують у режимах, де плавність керуючих сигналів є критичною умовою стійкості. Використання логічних перемикачів або дискретних правил зміни пріоритетів породжує розриви у функції винагороди та її частинних похідних, що може спричиняти коливальні режими, аеродинамічні перевантаження або втрату стабільності. Забезпечення глобальної диференційованості скаляризації залишається складною методологічною задачею.

3. Вразливість до експлуатації функції винагороди (reward hacking). Статично визначені функції винагороди створюють ризик їх формальної оптимізації без досягнення реальної операційної мети. У таких випадках агент знаходить стратегії, що максимізують числовий показник винагороди через прогалини в алгоритмах, не відповідаючи задуму розробника. У військовому контексті це може призвести до критичних наслідків, зокрема, до штучного підвищення показника «збереження боєзапасу» шляхом відмови від ураження легітимних цілей, що фактично суперечить оперативному завданню.

Для усунення зазначених недоліків пропонується метод стан-залежної нейромережевої скаляризації (SDNS), що трансформує процес скаляризації з пасивної арифметичної операції в активний інтелектуальний процес оцінки ситуації. Математична архітектура методу базується на тензорній конкатенації вектора локальних критеріїв, який доцільно записати у транспонованому вигляді:

$$Y(t) = |y_1(t), y_2(t), \dots, y_k(t), \dots, y_K(t)|^T \in \mathbb{R}^K, \quad (5)$$

та вектора поточного фазового стану системи:

$$X(t) = |x_1(t), x_2(t), \dots, x_n(t), .., \ x_N(t)|^T \in \mathbb{R}^N, \quad (6)$$

де $y_k(t)$ – значення критерію ефективності (2) в момент часу *t*, $k = (1, \dots, K)$;

$x_n(t)$ – значення змінної стану системи, в момент часу t, $n = (1, \dots, N)$;

$\mathbb{R}^K$ – простір критеріїв;

$\mathbb{R}^N$ – простір станів.

Вектор $X(t)$ може містити параметри кінематики, залишок пального, рівень радіоелектронного придушення, стан каналів зв'язку та інші телеметричні дані. Ці вектори конкатенуються у вхідний об'єднаний вектор ознак:

$$Z(t) = \begin{bmatrix} Y(t) \\ X(t) \end{bmatrix} = \begin{bmatrix} (Y(t)) & ,(X(t)) \end{bmatrix}^T \in \mathbb{R}^{K+N} \quad (7)$$

Тоді цільова функція ефективності (1), з урахуванням динамічних параметрів, трансформується у вигляд:

$$J_{SDNS}(t) = \Phi_\Theta\big(Z(t)\big), \quad (8)$$

де $\Phi_\Theta: \mathbb{R}^{K+N} \to \mathbb{R}$ – диференційований нейромережевий оператор;

Θ – множина її синаптичних ваг.

З урахуванням виразу (8) динамічна гранична норма заміщення (Marginal Rate of Substitution (далі – MRS)), наведена у виразі (2) між критеріями i та j визначається як відношення частинних похідних виходу мережі по відповідних входах:

$$MRS_{i,j}\big(X(t)\big) = \frac{\partial \Phi_\Theta}{\partial y_i} \cdot \left(\frac{\partial \Phi_\Theta}{\partial y_j}\right)^{-1}. \quad (9)$$

Завдяки безпосередній наявності вектора фазового стану $X(t)$ у вхідному шарі мережі та використанню математично гладких функцій активації (таких як Гаусова лінійна функція активації помилки (Gaussian Error Linear Unit (GELU)), функції активації Swish або Softplus, на відміну від розривних функцій типу випрямленої лінійної функція активації (Rectified Linear Unit, (ReLU)) або крокових функцій), зміна пріоритетів відбувається безперервно, миттєво та без розривів похідних [2]. Це забезпечує математично гладке формування керуючих сигналів.

Проте, інтеграція такої складної моделі з гігантським простором параметрів Θ у реальні автономні системи, особливо ті, що характеризуються летальним потенціалом, пов'язана з ризиками. Випадкова ініціалізація або некоректне застосування алгоритмів навчання може призвести до непередбачуваної, небезпечної поведінки автономного агента. Відповідно, процедура ідентифікації та навчання моделі SDNS вимагає іншого підходу. Тому пропонується підхід, який поєднує методи контрольованого навчання на історичних даних (Supervised Learning) та найсучасніші парадигми навчання з підкріпленням (далі – Meta-Reinforcement Learning), а також інтеграцію алгоритмів-наглядачів для коригування функції скаляризації в режимі реального часу.

Першим етапом у розгортанні архітектури SDNS є етап базової підготовки (Pre-training). У традиційному глибокому навчанні (Deep Learning) нейронні мережі часто ініціалізуються за допомогою стохастичних методів, таких як ініціалізація Ксав'є (Xavier initialization) або ініціалізація Хе (He initialization), які забезпечують оптимальну дисперсію активацій на початку тренування. Однак, у військовому контексті мережа SDNS не може бути ініціалізована випадково.

Причина цього криється в самій природі нелінійної скаляризації. Випадково згенеровані синаптичні ваги Θ продукуватимуть хаотичні, непередбачувані значення MRS. У перші ітерації функціонування така система генеруватиме керуючі сигнали, які можуть повністю ігнорувати критичні параметри безпеки, що неминуче призведе до катастрофічних рішень.

З огляду на зазначені ризики розглянемо етап базової підготовки (Pre-training). З метою уникнення формування нестабільних значень MRS на початку функціонування система відмовляється від класичних лінійних згорток зі статичними або випадково ініціалізованими вагами. Натомість SDNS реалізує диференційований нейромережевий оператор скаляризації, що забезпечує контекстно-залежну нелінійну адаптацію пріоритетів критеріїв.

Математично динамічна скаляризація визначається через диференційований нейромережевий оператор (8), що здійснює нелінійне відображення простору оцінок дій a у середовищі s за виразом:

$$R_\Theta(s,a) = \Phi_\Theta(f(s,a),s), \tag{9}$$

де – $f(s,a)$ – вектор локальних критеріїв ефективності дії a у стані середовища s;

Θ – множина параметрів нейронної мережі (ваги та зміщення).

Для ідентифікації параметрів Θ формується експертний датасет:

$$\mathcal{D} = \{(s_r, a_r, m_r^{exp})\}_{r=1}^{Q}, \tag{10}$$

де s_r – стан середовища для r-го рядка датасету;

a_r – дія, виконана у стані s_r;

m_r^{exp} – експертна оцінка ефективності виконаної дії;

Q – кількість записів у датасеті.

Ітеративне оновлення параметрів мережі Θ здійснюється завдяки використанню методів градієнтної оптимізації (наприклад, стохастичного градієнтного спуску) згідно з таким правилом:

$$\Theta_{\tau+1} = \Theta_\tau - \eta\nabla_\Theta L_{sup}(\Theta_\tau), \tag{11}$$

де τ – номер ітерації;

$\eta > 0$ – крок навчання;

$\nabla_\Theta L_{sup}(\Theta_\tau)$ – градієнт функції втрат.

Фундаментальним результатом цього етапу є початкова конфігурація оператора скаляризації, що апроксимує нормативну функцію оцінювання, узгоджену з військовими доктринами та принципами міжнародного гуманітарного права параметрів:

$$\Phi_\Theta \approx \Phi_{doctrine}, \tag{12}$$

Це гарантує стабільність MRS на початковому етапі експлуатації та виключає порушення критичних обмежень безпеки.

Після успішного завершення етапу базової підготовки (Pre-training), коли параметри нейромережевого оператора ініціалізовані до стану безпечної апроксимації, система набуває здатності адекватно оцінювати ситуацію. Однак, статичного навчання на історичних даних недостатньо для адаптації до високодинамічних та непередбачуваних умов реального середовища.

Тому логічним продовженням є перехід до другого етапу – адаптивного навчання в динамічному середовищі (Meta-Reinforcement Learning). На цьому етапі вводиться поняття активного агента (системи прийняття рішень), поведінка якого визначається стохастичною політикою:

$$\pi_\psi(a|s), \tag{12}$$

де ψ – параметри політики агента, що визначають ймовірність вибору дії a у стані середовища s.

Головним завданням агента є не просто максимізація скалярної нагороди, як у класичному навчанні з підкріпленням, а оптимізація очікуваного скаляризованого повернення. Цільова функція (8), яку доцільно максимізувати, набуває вигляду:

$$J_{SDNS}(\psi,\Theta) = E_{\pi_\psi}[\textstyle\sum_{\tau=0}^{T}\gamma^\tau\Phi_\Theta(f_t,s_t)], \tag{13}$$

де E_{π_ψ} – математичне сподівання за траєкторіями, згенерованими поточною політикою π_ψ;

T – горизонт планування (довжина епізоду);

$\gamma \in [0,1)$ – коефіцієнт дисконтування, що визначає вагомість майбутніх оцінок порівняно з поточними;

f_t – вектор значень критеріїв у момент часу t, що визначається залежністю $t = \tau\Delta t$;

Δt – крок дискретизації часу в горизонті планування;

$\Phi_\Theta(f_t,s_t)$ – динамічна винагорода, згенерована модулем SDNS на основі поточних параметрів скаляризації Θ.

s_t – n-вимірний вектор стану середовища у момент часу t.

Для оптимізації параметрів політики π_ψ застосовуються сучасні методи градієнта політики (Policy Gradient), зокрема, алгоритми сімейства PPO або MO-GRPO. Оновлення політики ґрунтується на обчисленні градієнта цільової функції (11):

$$\nabla_\psi J_{SDNS} = E\left[\nabla_\psi \log \pi_\psi\,(a_t|s_t)A_t^\Theta\right], \tag{14}$$

де A_t^Θ – модифікована функція переваги, що оцінює відносну ефективність дії a_t у стані s_t з урахуванням динамічної винагороди, сформованої оператором скаляризації SDNS.

На відміну від стандартних підходів функція переваги визначається завдяки безпосередньому використанню результатів роботи SDNS. Вона показує, наскільки обрана дія a_t є кращою за усереднену політику агента в стані s_t, з погляду поточної доктринальної скаляризації, заданої параметрами Θ.

Отже, на етапі метанавчання з підкріпленням (Meta-Reinforcement Learning) політика агента

безперервно адаптується, оптимізуючи очікуване скаляризоване повернення за множиною критеріїв, пріоритети яких формуються стан-залежним нейромережевим оператором SDNS. Це забезпечує узгоджене поєднання адаптивності, диференційованості та нормативної безпеки в процесі прийняття рішень.

Тому, з метою забезпечення загальної ефективності виконання місії параметри нейромережевого оператора скаляризації Θ підлягають періодичному оновленню у зовнішньому циклі (outer loop) мета – навчання. На відміну від агента, який керується поточними скаляризованими винагородами на кожному кроці τ, оптимізація модуля SDNS здійснюється на основі глобальної мета-цільової функції $J_{meta}(\Theta)$, що оцінює кінцевий результат усього епізоду (наприклад, кінцеву (термінальну) точність, факт виконання завдання або загальний рівень виживання системи).

Оскільки поведінка агента (його траєкторія g) безпосередньо залежить від заданих правил скаляризації, оновлення параметрів Θ відбувається вздовж напрямку мета-градієнта $\nabla_\Theta J_{meta}$. Відповідний ітераційний процес оновлення синаптичних ваг скаляризуючої мережі набуває вигляду:

$$\Theta_{new} = \Theta_{old} + \alpha_\Theta \nabla_\Theta E_{\tau\sim\pi_{\psi(\Theta)}}\left[R_{global}(\mathrm{g})\right], \quad (15)$$

де α_Θ – швидкість навчання на мета-рівні;

$R_{global}(\mathrm{g})$ – глобальна кумулятивна винагорода за траєкторію, згенеровану агентом, чия політика (12) була попередньо оптимізована у внутрішньому циклі за поточних параметрів Θ.

Практичне обчислення цього мета-градієнта $\nabla_\Theta J_{meta}$ для складних динамічних систем може здійснюватися за допомогою методів еволюційних стратегій (Evolution Strategies (ES)), що дає змогу стохастично оцінювати напрямок оновлення Θ і уникати обчислювально витратного розрахунку других похідних (матриці Гессе) крізь кроки градієнта політики (11).

Отже, методологічні засади SDNS зводяться до інтеграції класичної теорії багатокритеріальної оптимізації з передовими парадигмами машинного навчання, що перетворює скаляризацію зі статичної обчислювальної процедури на безперервний інтелектуальний процес оцінювання ситуації. Практично ці засади реалізуються через побудову багаторівневого контуру оптимізації: від попереднього формування безпечних базових пріоритетів на основі експертних даних (Supervised pre-training), через швидку реактивну адаптацію виконавчої політики агента (Inner-loop RL), і аж до стратегічної еволюції самої функції динамічної винагороди (Outer-loop Meta-RL). Завдяки такій синергії підходів, що функціонує під жорстким контролем алгоритмів-наглядачів, метод вирішує фундаментальну проблему управління у непередбачуваних умовах. Він наділяє систему здатністю автономно, гнучко та математично обґрунтовано обмінювати локальні критерії заради безумовного виконання глобального завдання місії.

З метою систематизації отриманих результатів та формалізованого порівняння запропонованого підходу з існуючими методами скаляризації доцільно узагальнити їх основні математичні та функціональні характеристики у вигляді порівняльної матриці.

У таблиці 1 наведено порівняльний аналіз трьох класів операторів скаляризації: лінійної зваженої суми, гладкої мінімаксної згортки Чебишова (STCH) та запропонованої стан-залежної нейромережевої скаляризації (SDNS):

Таблиця 1

Порівняльний аналіз математичних властивостей операторів скаляризації

Характеристика оператора скаляризації	Лінійний оператор скаляризації	Нелінійний аналітичний оператор (STCH)	Стан-залежний нейромережевий оператор (SDNS)
Математична природа агрегації	Алгебраїчна, лінійна	Аналітична, нелінійна	Нейромережева, глибоко нелінійна
Здатність знаходити рішення у неопуклих ділянках фронту Парето	Ні (обмежена опуклістю фронту)	Так (через мінімаксну конструкцію)	Так (завдяки універсальній апроксимації)
Гладкість та диференційованість	Так	Так (через згладжувальні функції)	Так (за умови використання гладких активацій, напр. GELU, Swish)
Адаптивність пріоритетів до фазового стану системи (X_t)	Ні (потребує зовнішніх правил)	Ні (параметри преференцій фіксовані)	Так (фазовий стан інтегрований у простір ознак)
Характер граничної норми заміщення (MRS)	Фіксована константа $\left(\frac{w_j}{w_i}\right)$	Змінна, аналітично визначена	Динамічна, стан-залежна, еволюціонує у процесі мета-навчання

Порівняння здійснено за критеріями математичної природи агрегації, здатності працювати з неопуклими фронтами Парето, диференційованості, адаптивності до фазового стану та характеру граничної норми заміщення. Це дає змогу обґрунтовано продемонструвати методологічні переваги SDNS

Результати аналізу табл. 1 свідчить, що SDNS поєднує переваги аналітичної гладкості з адаптивністю та здатністю працювати з неопуклими структурами множини Парето, що принципово розширює можливості багатокритеріального управління динамічними системами військового призначення.

Висновки

Традиційні лінійні методи згортки є обмеженими у складних тактичних сценаріях через геометричну нездатність охоплювати неопуклі ділянки фронту Парето, а також через фіксовану граничну норму заміщення, що не відображає контекстно-залежну цінність ресурсів у динамічному середовищі.

Запропоновані методологічні засади стан-залежної нейромережевої скаляризації (State-Dependent Neural Scalarization, (SDNS)) усувають зазначені обмеження завдяки інтеграції вектора фазового стану безпосередньо в оператор скаляризації критеріїв. Це дає змогу реалізувати неперервну, нелінійну та диференційовану зміну пріоритетів, що гарантує математичну гладкість керуючих сигналів і стійкість контуру управління. Інтеграція нейромережевих моделей у бойові автономні системи вимагає суворого контролю, оскільки випадкова ініціалізація параметрів здатна призвести до ігнорування критеріїв безпеки та катастрофічних рішень.

З метою уникнення цього, запропоновано двоетапне розгортання архітектури стан-залежної нейромережевої скаляризації (SDNS): по-перше застосовується інтерактивне навчання диференційованого нейромережевого оператору (8) на експертних датасетах для мінімізації емпіричного ризику та узгодження оцінок з нормами міжнародного гуманітарного права. По-друге, оптимізуються політики методом градієнта політики в межах метанавчання з підкріпленням (Meta-Reinforcement Learning) для максимізації очікуваного скаляризованого повернення.

Перспективи подальших досліджень. Незважаючи на потужність архітектури стан-залежної нейромережевої скаляризації (SDNS), критично важливою перспективою подальших досліджень є розроблення комплексної моделі середовища для навчання з підкріпленням (Reinforcement Learning (RL)). Оскільки другий етап розгортання архітектури стан-залежної нейромережевої скаляризації (SDNS) передбачає взаємодію активного агента з високодинамічними умовами, існує нагальна потреба у створенні високоточної симуляційної платформи. Це середовище має адекватно генерувати вектори поточного фазового стану, зокрема, динамічні параметри, рівень радіоелектронного придушення, стан каналів зв'язку та інші телеметричні дані тощо. Крім того, практична реалізація та інтеграція алгоритмів-наглядачів передбачатиме безперервний моніторинг та екстрене коригування функції скаляризації у реальному часі.

Конфлікт інтересів. Конфлікти інтересів, що впливають на результати дослідження, відсутні.

Фінансування. Фінансування дослідження не здійснювалося.

Доступність даних. Дослідження виконано з використанням виключно відкритих даних, доступних у публічних джерелах.

Використання засобів штучного інтелекту (далі – ШІ). Під час підготовки статті авторами застосовувалися засоби штучного інтелекту. Зокрема, інструменти ШІ використовувалися для інформаційного пошуку наукових джерел, попереднього аналізу, перевірки формалізації окремих математичних положень, технічного редагування тексту, а також для перекладу анотації та окремих фрагментів статті англійською мовою. Застосування засобів ШІ мало виключно допоміжний характер і не передбачало автоматичного формування наукових висновків без участі авторів. Усі результати, отримані із використанням інструментів ШІ, були критично проаналізовані та верифіковані авторами. Використання засобів ШІ не призвело до порушення авторських прав та норм академічної етики. Згенерований або оброблений за допомогою ШІ контент був перевірений, відредагований та відповідає фактичним даним і змісту проведеного дослідження.

Список бібліографічних посилань

1. Artificial intelligence in the military domain and its implications for international peace and security: the UN General Assembly on 24 December 2024 № 79/239. URL https://docs-library.unoda.org/General_Assembly_First_Committee_-Eightieth_session_%282025%29/79-239-Ukraine-EN.pdf?utm_source=chatgpt.com (дата звернення: 14.02.2026). **2. Miettinen K. M.** Nonlinear Multiobjective Optimization. Boston: Kluwer Academic Publishers, 1999. 324 p. **3. Deb K.** Multi-Objective Optimization Using Evolutionary Algorithms. Boston: Springer, 2001. 497 p. **4. Sutton R. S., Barto A. G.** Reinforcement Learning: An Introduction. 2nd ed, Cambridge, MA: The MIT Press, 2018. 552 p. **5. Про затвердження плану заходів з реалізації Концепції розвитку штучного інтелекту в Україні на 2021–2024 роки.** Постанова КМ України від 12.05.2021 № 438-р. URL: https://zakon.rada.gov.ua/laws/main/438-2021-%D1%80?utm_source=chatgpt.com#Text (дата звернення: 14.02.2026). **6. Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy.** Washington, D.C.: U.S. Department of State, 2023. URL: https://www.state.gov/wp-content/uploads/2023/10/Latest-Version-Political-Declaration-on-Responsible-Military-Use-of-AI-and-Autonomy.pdf (дата звернення: 25.02.2026). **7. Калачова В. В., Ткачук С. С., Меренті Є. О., Третяк Д. В.** Багатокритеріальний синтез організаційної структури білінгової інформаційної системи методом аналізу ієрархій. *Системи обробки інформації*. 2020. № 2(161). С. 22–28. DOI: https://doi.org/10.30748/soi.2020.161.03. **8. Крайнов В., Грозовський Р., Кравчук А.** Методика оцінки якості інформаційно-аналітичного забезпечення роботи автоматизованих інформаційних систем органів управління військового призначення. *Сучасні інформаційні технології у сфері безпеки та оборони*. 2019. № 3(36). С. 69–74. DOI: https://doi.org/10.33099/2311-7249/2019-36-3-69-74.

9. Кучук А. О. Моделі штучного інтелекту для управління та оптимізації peer-to-peer мереж. *Радіоелектроніка та молодь у XXI столітті:* матеріали 27-го Міжнар. молодіж. форуму, Харків, 10–12 травня 2023 р. ХНУРЕ, Харків, 2023. Т. 6. Ч. 1. С. 27–28. **10. Lin X., Zhang X., Yang Z., Liu F., Wang Z., Zhang Q.** Smooth Tchebycheff Scalarization for Multi-Objective Optimization. *Proceedings of the 41st International Conference on Machine Learning (ICML 2024)*. New York, NY, USA: ACM, 2024. DOI: https://doi.org/10.5555/3692070.3693297. **11. Peng N., Tian M., Fain B.** Multi-objective Reinforcement Learning with Nonlinear Preferences: Provable Approximation for Maximizing Expected Scalarized Return. *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025)*. Detroit, MI, USA: International Foundation for Autonomous Agents and Multiagent Systems, 2025. URL: https://www.ifaamas.org/Proceedings/aamas2025/pdfs/p1632.pdf (дата звернення: 14.02.2026). **12. Ichihara Y., Jinnai Y., Morimura T., Sakamoto M., Mitsuhashi R., Uchibe E.** MO-GRPO: Mitigating Reward Hacking of Group Relative Policy Optimization on Multi-Objective Problems. 2025. (Preprint. arXiv:2509.22047). URL: https://arxiv.org/abs/2509.22047 (дата звернення: 25.02.2026). **13. Cybenko G**. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems.* 1989. Vol. 2. P. 303–314. DOI: https://doi.org/10.1007/BF02551274. **14. Hornik K., Stinchcombe M., White H.** Multilayer feedforward networks are universal approximators. *Neural Networks*. 1989. Vol. 2. No. 5. P. 359–366. DOI: https://doi.org/10.1016/0893-6080(89)90020-8.

METHODOLOGICAL FOUNDATIONS OF NEURAL NETWORK SCALARIZATION FOR MULTI-CRITERION CONTROL PROBLEMS IN DYNAMIC MILITARY SYSTEMS

***MYKOLAICHUK Roman*,** Doctor of Technical Sciences, Associate Professor, National Defence University of Ukraine, Kyiv, Ukraine, https://orcid.org/0000-0001-5349-4487

***LYFAR Olena*,** National Defence University of Ukraine, Kyiv, Ukraine, https://orcid.org/0009-0004-4894-2184

Formulation of the problem in general. *Modern military artificial intelligence systems operate in highly dynamic environments characterised by conflicting objectives such as mission efficiency, safety, resource preservation, and compliance with legal and ethical constraints. Classical linear scalarization methods are unable to adequately handle non-convex Pareto fronts and context-dependent priority shifts in real-time decision-making.* ***Purpose of the article.*** *The purpose of the article is to develop the methodological foundations of State-Dependent Neural Scalarization as a universal, differentiable operator for transforming vector-multiple-criteria objective functions into scalar form for dynamic military control systems.*

Research methods. *The study applies methods of systems analysis, multi-criteria optimisation theory, Pareto analysis, functional analysis, tensor algebra, deep learning, and Multi-Objective Reinforcement Learning. Gradient-based optimisation and differentiability analysis of nonlinear aggregation operators are used to ensure smooth adaptive control.*

Literature review. *Existing scalarization approaches, including linear weighted sums and Smooth Tchebycheff scalarization, were analysed. Although nonlinear methods improve coverage of non-convex Pareto regions, they remain analytically rigid and do not incorporate the system's dynamic phase state into the aggregation mechanism.*

Research results. *A formal mathematical architecture of State-Dependent Neural Scalarization is proposed based on the tensor concatenation of the criteria vector and the system phase-state vector. A differentiable neural aggregation operator is developed, enabling dynamic adjustment of priorities through state-dependent marginal rates of substitution. A two-stage learning procedure combining supervised pre-training and Meta-Reinforcement Learning is formulated.*

Research novelty. *For the first time, a state-dependent neural scalarization methodology is formalised as a differentiable operator parameterised by the system phase state, enabling adaptive nonlinear aggregation of criteria in dynamic military environments.*

The theoretical and practical significance *of the results presented in this paper lies in the formalisation of state-dependent neural network scalarization as a differentiable operator for criteria aggregation in multi-criteria military dynamic systems. The proposed method provides a mathematically consistent transformation of a vector-valued objective function into a scalar form while accounting for the phase state of the controlled object. This extends the methodology of multi-criteria optimisation and establishes a foundation for integrating neural network methods into adaptive control loops. The practical significance lies in the possibility of directly implementing the State-Dependent Neural Scalarization method in Command and Control (C2) systems, autonomous unmanned platforms, electronic warfare countermeasure systems, and autonomous cyber defence architectures. The proposed approach enables adaptive real-time switching of priorities between mission effectiveness, safety constraints, and resource limitations. Furthermore, it enhances the robustness of reinforcement learning algorithms under dynamically changing tactical conditions and reduces the risk of undesirable or misaligned behaviour of autonomous agents.*

Conclusions and future work. *State-Dependent Neural Scalarization overcomes geometric and adaptive limitations of classical scalarization methods. Future research should focus on developing high-fidelity simulation environments for reinforcement learning and on integrating supervisory safety algorithms to correct the scalarization function in real time.*

Keywords: *artificial intelligence, machine learning, artificial neural networks, reinforcement learning, performance evaluation, multi-criteria optimisation, neural network scalarization, control theory, data processing, management of dynamic military command and control systems.*

References

1. Artificial intelligence in the military domain and its implications for international peace and security: the UN General Assembly on 24 December 2024 № 79/239 [online]. Available at: https://docs-library.unoda.org/General_Assembly_First_Committee_-Eightieth_session_%282025%29/79-239-Ukraine-EN.pdf?utm_source=chatgpt.com [Accessed: 14 February 2026]. **2. Miettinen, K. M., (1999). Nonlinear Multiobjective Optimisation**. Boston: Kluwer Academic Publishers, 1999. **3. Deb, K.,** *Multi-Objective Optimisation Using Evolutionary Algorithms.* Boston: Springer, 2001. **4. Sutton, R. S., Barto, A. G**. *Reinforcement Learning: An Introduction.* 2nd ed., Cambridge, MA: The MIT Press, 2018. **5. On Approval of the Action Plan for the Implementation of the Concept for the Development of Artificial Intelligence in Ukraine for 2021–2024.** Resolution of the Cabinet of Ministers of Ukraine No. 438-r dated 12 May 2021. Available at: https://zakon.rada.gov.ua/laws/main/438-2021-%D1%80#Text [Accessed: 14 February 2026]. **6. Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy.** Washington, D.C.: U.S. Department of State, 2023. Available at: https://www.state.gov/wp-content/uploads/2023/10/Latest-Version-Political-Declaration-on-Responsible-Military-Use-of-AI-and-Autonomy.pdf [Accessed: 14 February 2026]. **7. Kalachova, V.V., Tkachuk, S.S., Merenti, Ye.O., Tretiak, D.V.** 2020 Multicriteria Synthesis of the Organisational Structure of a Billing Information System Using the Analytic Hierarchy Process. *Information Processing Systems*, 2(161), 22-28. DOI: https://doi.org/10.30748/soi.2020.161.03. **8. Krainov, V., Hrozovskyi, R., Kravchuk, A.,** (2019). Methodology for Assessing the Quality of Information and Analytical Support of Automated Information Systems for Military Management Authorities. *Modern Information Technologies in the Sphere of Security and Defence*, 36, 3, 69-74. DOI: https://doi.org/10.33099/2311-7249/2019-36-3-69-74. **9. Kuchuk, A.O.,** (2023). Artificial Intelligence Models for Management and Optimisation of Peer-to-Peer Networks. In: *Radio Electronics and Youth in the 21st Century: Proceedings of the 27th International Youth Forum*, Kharkiv, May 10-12, 2023. Kharkiv: KhNURE, 6, 1, 27-28. **10. Lin, X., Zhang, X., Yang, Z., Liu, F., Wang, Z., & Zhang, Q.,** (2024). Smooth Tchebycheff Scalarization for Multi-Objective Optimisation. *Proceedings of the 41st International Conference on Machine Learning (ICML 2024)*. New York, NY, USA: ACM. DOI: https://doi.org/10.5555/3692070.3693297. **11. Peng, N., Tian, M., & Fain, B.,** (2025). Multi-objective Reinforcement Learning with Nonlinear Preferences: Provable Approximation for Maximising Expected Scalarized Return. *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025)*. Detroit, MI, USA: International Foundation for Autonomous Agents and Multiagent Systems. Available at: https://www.ifaamas.org/Proceedings/aamas2025/pdfs/p1632.pdf [Accessed: 14 February 2026]. **12. Ichihara, Y., Jinnai, Y., Morimura, T., Sakamoto, M., Mitsuhashi, R., Uchibe, E.** MO-GRPO: Mitigating Reward Hacking of Group Relative Policy Optimisation on Multi-Objective Problems. 2025. (Preprint. arXiv:2509.22047). Available at: https://arxiv.org/abs/2509.22047 [Accessed: 25 February 2026]. **13. Cybenko, G**., (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems,* 2, 303-314. DOI: https://doi.org/10.1007/BF02551274. **14. Hornik, K., Stinchcombe, M., White, H.,** (1989). Multilayer feedforward networks are universal approximators. *Neural Networks,* 2, 5, 359-366. DOI: https://doi.org/10.1016/0893-6080(89)90020-8.

Рукопис надійшов до редакції *28.02.2026*
Рукопис прийнято до друку після рецензування *06.04.2026*
Дата публікації *30.04.2026*