

**ВІДГУК**  
на дисертаційну роботу  
Стрюка Олександра Сергійовича

на тему «Оптимізація генеративних змагальних нейронних мереж в умовах  
апаратно-параметричних обмежень»,

представлену на здобуття ступеня доктора філософії  
в галузі знань 12 Інформаційні технології  
за спеціальністю 123 Комп'ютерна інженерія

**Актуальність теми дисертації.**

Сучасні системи штучного інтелекту (ШІ) відіграють ключову роль в автоматизації, контролі стану об'єктів критичної інфраструктури та розв'язанні складних аналітичних задач. Однак ефективність цих систем значною мірою залежить від обчислювальних ресурсів, оскільки застосування штучних нейронних мереж та моделей генеративного ШІ традиційно спирається на спеціалізовані обчислювальні машини з потужними дискретними графічними картами або на сервіси хмарних обчислень.

Традиційні підходи до розгортання генеративних змагальних штучних нейронних мереж (ГЗМ) часто виявляються недостатньо гнучкими для автономної роботи в режимі реального часу або в умовах, що вимагають суворої конфіденційності даних, через брак обчислювальних потужностей, затримки передачі інформації та ризики кібератак. У сучасних умовах зростає потреба у вдосконаленні методів проектування та навчання ГЗМ безпосередньо на мікрокомп'ютерах та кордонних пристроях (Edge AI).

Використання новітніх підходів до оптимізації навчання систем генеративного ШІ дозволяє значно знизити обчислювальне навантаження, подолати проблеми нестабільності навчання (зникнення градієнтів, колапс моди) та підвищити ефективність роботи інтелектуальних систем.

Додатково, швидкий розвиток Інтернету речей (IoT) відкриває нові можливості для інтелектуального аналізу даних. Проте інтеграція ресурсномістких генеративних архітектур в існуючі вбудовані системи потребує розробки стандартизованих підходів до їх оптимізації з урахуванням апаратно-параметричних обмежень.

У цьому контексті актуальною є розробка нових методів оптимізації навчання та структурної реконфігурації ГЗМ для забезпечення їх стабільної конвергенції та функціонування на кордонних пристроях. Традиційні алгоритми оптимізації не завжди враховують дефіцит оперативної пам'яті та

ліміти енергоспоживання, що зменшує їхню ефективність у реальних науково-практичних задачах.

Таким чином, удосконалення методів оптимізації ГЗМ в умовах апаратно-параметричних обмежень є актуальним науково-практичним завданням. Запропоновані в дисертаційному дослідженні методи та моделі мають суттєве значення для підвищення автономності інтелектуальних систем при розгортанні на кордонних та мікрокомп'ютерних пристроях при виконанні задач комп'ютерного зору, біометрії, виявлення аномалій, а також розширення можливостей IoT-платформ.

**Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни.**

Наукова новизна результатів дисертаційного дослідження полягає в наступному:

1) **вперше розроблено** математичну модель каскадної оптимізації ГЗМ, яка, на відміну від існуючих, базується на ієрархічній декомпозиції простору гіперпараметрів та врахуванні адаптивної динаміки навчання, що дозволяє забезпечити необхідну точність та швидкодію функціонування ГЗМ в умовах апаратно-параметричних обмежень кордонних пристроїв;

2) **вперше розроблено** мультифазовий метод оптимізації навчання ГЗМ, який, на відміну від існуючих, базується на багаторівневому механізмі адаптивної конвергенції, що дозволяє запобігати колапсу моди та зникненню градієнтів функцій втрат без підвищення обчислювальної складності процедури навчання;

3) **удосконалено** механізм адаптації ГЗМ, який ґрунтується на гібридному комбінуванні каскадного та мультифазового методів в поєднанні з апаратом нечіткої логіки, що дозволяє комплексно підвищити ефективність навчання та якість генерації штучних даних, зокрема знизити функцію втрат генератора у 3,5 рази, прискорити збіжність ГЗМ в 7,6 разів та покращити метрику FID у 2,3 рази;

4) **набув подальшого розвитку** програмно-апаратний метод реалізації повного циклу функціонування ГЗМ на кордонних пристроях, який базується на інтеграції квантованого навчання та апаратно-орієнтованої каскадної оптимізації, що забезпечує реалізацію реконфігурованих архітектур зі зменшенням розміру імітаційної моделі в 3,9 рази та прискоренням процесу інференсу в 3,2 рази.

Наукові положення, викладені в дисертаційній роботі є достатньо обґрунтованими, а отримані результати узгоджуються з вже відомими та є їх подальшим розвитком. Здобувач здійснив глибокий аналіз та систематизацію наукових результатів, використовуючи ретельно обґрунтовані методи дослідження. Достовірність отриманих результатів підкріплюється

моделюванням та експериментальними дослідженнями на базі мікрокомп'ютерної платформи Raspberry Pi 5.

Отже, в дисертаційній роботі поставлене наукове завдання виконано повністю, здобувач повною мірою оволодів методологією наукової діяльності.

**Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності.**

За своїм змістом дисертаційна робота здобувача Стрюка О. С. повністю відповідає Стандарту вищої освіти зі спеціальності та напрямкам досліджень відповідно до освітньої програми спеціальності 123 Комп'ютерна інженерія.

Дисертаційна робота є завершеною науковою працею і свідчить про наявність особистого внеску здобувача у науковий напрям «Інформаційні технології».

Розглянувши звіт подібності за результатами перевірки дисертаційної роботи на текстові співпадіння, можна зробити висновок, що дисертаційна робота Стрюка Олександра Сергійовича є результатом самостійних досліджень здобувача і не містить елементів фальсифікації, компіляції, фабрикації, плагіату та запозичень. Використані ідеї, результати і тексти інших авторів мають належні посилання на відповідне джерело.

**Мова та стиль викладення результатів.**

Дисертаційна робота написана українською мовою. Результати досліджень викладені послідовно та лаконічно, що дозволяє читачеві легко розібратися у методології та отриманих висновках. Автор використовує доступний стиль мовлення, що сприяє зрозумінню матеріалу. Використання загальноприйнятої термінології забезпечує консистентність і точність викладення інформації, що є важливим фактором для наукових досліджень.

У вступі розкрито актуальність теми дослідження, приведено зв'язок роботи з науковими програмами, планами, темами, наведено мету і завдання дослідження, методи дослідження, наукову новизну та практичне значення отриманих результатів, особистий внесок здобувача, апробацію результатів дисертації, кількість публікацій та обсяг роботи, що відповідає загальноприйнятій структурі дисертаційної роботи.

**Скорочений опис розділів дисертації.**

Перший розділ дисертації присвячений аналізу літературних джерел в області теоретичних основ проєктування та оптимізації ГЗМ, прикладного використання ГЗМ з реалізацією їх на мікропроцесорних і кордонних пристроях та мікрокомп'ютерах, перспектив їх застосування в різних областях людської діяльності. В результаті аналізу сучасного стану розвитку ГЗМ встановлено, що врахування особливостей проєктування ГЗМ на кордонних пристроях та їх апаратно-параметричних обмежень (АПО) може здійснюватися шляхом (а) структурної реконфігурації архітектур ГЗМ, (б)

розробки нових та модифікованих методів оптимізації та їх гібридної комбінації для процесів нейромережевого навчання та (в) шляхом апаратно-орієнтованої адаптації обчислювальних процесів і кордонних пристроїв.

У другому розділі проаналізовано математичні моделі, функції втрат та алгоритми навчання ГЗМ, а також методи досягнення конвергенції. Обґрунтовано стратегії стабілізації процесів навчання ГЗМ на основі модифікації функцій втрат та активації, пакетної нормалізації, транспонованої згортки, підрізки ваг та градієнтного штрафу з фокусом на оптимізацію гіперпараметрів та адаптацію архітектур для забезпечення необхідної продуктивності, швидкодії та точності в умовах АПО кордонних пристроїв. Окрему увагу приділено аналізу недоліків існуючих ГЗМ, в тому числі з реалізацією на кордонних пристроях, та методів їх проектування. Детально висвітлені особливості процесів навчання ГЗМ з виникненням колапсу моди, конвергенції мереж генератора і дискримінатора, зникненням градієнтів функцій втрат та перенавчанням. Підкреслено потенціал структурно-параметричної оптимізації ГЗМ для адаптації їх математичних моделей, модифікації алгоритмів навчання та організації обчислювальних процесів на кордонних пристроях в умовах обмежень їх процесорної потужності, обсягів пам'яті і швидкодії.

Третій розділ дисертації присвячений стратегіям та методам оптимізації ГЗМ в умовах АПО з забезпеченням стабільності процесів навчання і підвищення якості генерації. Запропоновано каскадний метод оптимізації і механізм структурно-параметричного навчання ГЗМ, що забезпечує узгодження архітектури, гіперпараметрів і темпу навчання, мінімізуючи вплив експертних помилок на кінцеву ефективність моделі та її збіжність. Додатково розроблено метод мультифазової оптимізації, що розділяє процес навчання на декілька етапів, націлених на усунення окремих недоліків ГЗМ. Для підвищення ефективності та стабілізації навчання ГЗМ запропоновано механізм гібридизації каскадного та мультифазового методів.

Четвертий розділ дисертації присвячений імітаційному моделюванню та експериментальній валідації розроблених методів оптимізації для архітектур на основі стандартної ГЗМ та її модифікацій з різними функціями втрат (бінарної перехресної ентропії, втрат Васерштейна зі штрафом за градієнт, перцептивних та змагальних втрат, втрат контенту та середньої абсолютної похибки, втрата МНК та ін.). Детально описано результати експериментів (реалізація ГЗМ на мові Python 3 та TensorFlow, Keras, PyTorch), проведених для перевірки ефективності розроблених методів (у різних контекстах застосування ГЗМ) на датасетах MNIST, SOCOFing та BIRDS 400, зокрема для задач біометрії, покращення роздільної здатності та виявлення аномалій з використанням комбінованих функцій втрат. В ході проведених експериментів

вдалось забезпечити зниження функції втрат генератора (з 9,16 до 2,57), уникнути перенавчання дискримінатора та стабілізувати його точність (на рівні 92%), прискорити збіжність нейромережі (з 1300 до 170 епох), покращити метрику FID (з 45,9 до 19,4), а також досягти точності виявлення аномалій (AUC = 0,92) із забезпеченням повноти Recall (на рівні 1.0).

П'ятий розділ присвячено експериментальній верифікації адаптованих архітектур ГЗМ на кордонних пристроях (концепції On-Device Edge AI) для їх функціонування в умовах АПО. Серед спектру сучасних платформ для кордонних обчислень (таких як NVIDIA Jetson, Google Coral Dev Board, Intel Neural Compute Stick або FPGA-рішення) обґрунтовано в якості експериментальної платформи обрання одноплатного мікрокомп'ютера Raspberry Pi 5 (процесор Broadcom BCM2712, архітектура ARM Cortex-A76, 8 ГБ RAM). Raspberry Pi 5 не має спеціалізованих тензорних ядер, але його перевагами є низьке енергоспоживання, масо-габаритна компактність, універсальність до застосування в Інтернеті речей та низька вартість. Детально описано ключовий етап реалізації динамічного квантування ваг моделі у цілочисельний формат INT8 (засобами бібліотеки torch.quantization) з формату з плаваючою комою FP32. Це забезпечило критичне зниження ресурсомісткості ГЗМ: розмір моделі зменшено з 2,14 МБ до 0,55 МБ, що дозволяє розміщувати модель у швидкій кеш-пам'яті процесора. Профілювання інференсу (PyTorch Profiler) підтвердило прискорення часу відгуку до 0,61 мс порівняно з базовою FP32-версією (1,95 мс), при збереженні прийнятної якості генерації зразків. Використання каскадного та мультифазового методів оптимізації також дозволило досягти структурної узгодженості та покращення метричних показників коректності відтворення при генерації рукописного тексту MNIST, як тестового типу даних. Отримані і детально представлені в розділі результати доводять, що оптимізовані ГЗМ здатні працювати автономно в режимі реального часу на кордонних пристроях, усуваючи залежність від хмарних обчислень.

### **Оприлюднення результатів дисертаційної роботи.**

Загальні положення дослідження висвітлено в 17 наукових публікаціях, з них: 3 статті в міжнародних фахових журналах, 2 статті у наукових фахових виданнях України (які входять до переліку МОН України), 2 розділи в англійськомовних монографіях, 1 колективна монографія, 6 матеріалів доповідей на міжнародних науково-практичних конференціях (IDAACS'2023, DESSERT'2023, АПІТ'2020, АІСТ'2021, СМІС'2025 та ін.), 3 тез доповідей на Всеукраїнських науково-практичних конференціях.

Здобувач продемонстрував високий науковий рівень у своїх публікаціях, який відображається в їхній актуальності, методичності та оригінальності, про що свідчить наявність 11 публікацій у базі даних Scopus. Здобувач

дотримувався принципів академічної доброчесності, представляючи результати своїх досліджень з чітким викладом джерел та відсутністю плагіату. Таким чином, наукові результати, описані в дисертаційній роботі, повністю висвітлені у наукових публікаціях здобувача.

### **Недоліки та зауваження до дисертаційної роботи.**

У роботі представлено використання ГЗМ для виявлення аномалій, однак показник точності (precision) для аномальних зразків становить 0,12, що вказує на хибнопозитивні спрацювання. Доцільно було б доповнити дослідження порівняльним аналізом із результатами класичних (baseline) алгоритмів детекції аномалій.

У другому розділі дисертації наведено занадто розгорнутий опис загальновідомих цільових мінімаксних функцій ГЗМ, що надає йому дещо оглядового характеру. Необхідно було б детальніше сфокусуватися безпосередньо на апаратно-орієнтованих обмеженнях цільових функцій.

При описі завантаженості хмарного середовища під час навчання автор вказує навантаження CPU на рівні 150%. Це пояснюється підключенням GPU, однак у тексті бракує формалізованого математичного опису механізму перерозподілу потоків і масштабування обчислень.

Також доцільним було б додати формалізацію математичного опису процесу квантування ГЗМ, що підсилило б експериментальну частину п'ятого розділу дисертації.

Підводячи підсумок критичного розгляду матеріалів дисертаційної роботи, зазначу, що більшість зауважень, як видно із відгуку, торкається формальної чи технічної сторони оформлення роботи. Решта були сформульовані з метою поглиблення уяви здобувача про сутність обчислювальних процесів та звернення уваги на певні особливості обраної сфери дослідження.

Вважаю, що висловлені зауваження не є визначальними і не зменшують загальну наукову новизну та практичну значимість результатів та не впливають на позитивну оцінку дисертаційної роботи.

### **Висновок про дисертаційну роботу.**

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Стрюка Олександра Сергійовича на тему «Оптимізація генеративних змагальних нейронних мереж в умовах апаратно-параметричних обмежень» виконана на високому науковому рівні, не порушує принципів академічної доброчесності та є закінченим науковим дослідженням, сукупність теоретичних та практичних результатів якого розв'язує наукове завдання, що має істотне значення для комп'ютерної інженерії та штучного інтелекту.

Дисертаційна робота за актуальністю, практичною цінністю та науковою новизною повністю відповідає вимогам чинного законодавства України, що

передбачені в пунктах 6-9 «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого постановою Кабінету Міністрів України від 12 січня 2022 р. № 44.

Здобувач Стрюк Олександр Сергійович заслуговує на присудження ступеня доктора філософії в галузі знань 12 Інформаційні технології за спеціальністю 123 Комп'ютерна інженерія.

**Офіційний опонент:**

професор кафедри штучного інтелекту  
Харківського національного університету  
радіоелектроніки, Заслужений діяч науки і техніки України,  
доктор технічних наук,  
професор



Євгеній БОДЯНСЬКИЙ