**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.**

Безперервне ускладнення задач моніторингу та управління в сучасних кіберфізичних комплексах вимагає переходу до децентралізованих методів обробки інформації. Розв'язання цього класу задач дедалі частіше покладається на нейромережевий апарат, зокрема на архітектури генеративного штучного інтелекту. Проте з інженерної точки зору їх практичне застосування суттєво обмежується надмірною ресурсомісткістю сучасних архітектур.

Класична орієнтація на високопродуктивні хмарні кластери чи потужні графічні співпроцесори є неприйнятною для автономних пристроїв. Це зумовлено високими затримками на інтерфейсах передачі даних, критичною вразливістю каналів зв'язку до зовнішніх втручань та неможливістю гарантувати відмовостійкість у режимі реального часу.

Таким чином, виникає гостра науково-технічна суперечність: з одного боку, існує нагальна потреба в інтелектуалізації кінцевих вузлів Інтернету речей (IoT), а з іншого — наявна жорстка нестача обчислювальної потужності, обмеженість кеш-пам'яті та енергетичних ресурсів вбудованих мікрокомп'ютерних систем. Пряма імплементація складних генеративних змагальних нейромереж (ГЗМ) у такі апаратні платформи неминуче призводить до деградації їхньої продуктивності та апаратного троттлінгу.

У зв'язку з цим, перенесення обчислень на кордонні пристрої (концепція Edge AI) вимагає не простої адаптації програмного коду, а розробки принципово нових методів апаратно-програмного співпроєктування. Головне завдання полягає у досягненні стабільної алгоритмічної збіжності та мінімізації обчислювального навантаження без втрати якості генерації і аналізу даних, що вимагає специфічної структурної реконфігурації нейромереж.

З огляду на вищезазначене, дисертаційна робота Стрюка О.С., яка присвячена вирішенню проблеми оптимізації ГЗМ з чітким урахуванням апаратно-параметричних обмежень мікрокомп'ютерних платформ, є своєчасною та безперечно актуальною. Отримані в дисертаційній роботі результати відкривають шлях до створення нового класу автономних, стійких

до відмов вбудованих систем для задач технічного зору, виявлення аномалій та біометричної ідентифікації, що має вагоме значення для розвитку комп'ютерної інженерії та систем штучного інтелекту.

**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.

Основні наукові положення, представлені в дисертації, характеризуються належним рівнем обґрунтованості, а здобуті результати добре узгоджуються з існуючими науковими розробками, забезпечуючи їх подальший розвиток. Автором проведено ґрунтовний аналіз і систематизацію наукових результатів із застосуванням належних методів дослідження. Достовірність запропонованих рішень підтверджується результатами імітаційного моделювання та експериментами на базі мікрокомп'ютерної платформи Raspberry Pi 5.

Таким чином, сформульоване наукове завдання вирішене в повному обсязі, що свідчить про глибоке оволодіння здобувачем методологією науково-дослідної роботи.

**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».

На підставі аналізу звіту про перевірку тексту рукопису на наявність текстових збігів, підтверджено високий рівень оригінальності дисертаційної роботи Стрюка О.С. Представлене дослідження виконано здобувачем цілком самостійно. У тексті не виявлено ознак академічного плагіату, фабрикації даних чи неправомірної компіляції. Всі запозичені з інших праць ідеї, теоретичні концепції та результати супроводжуються коректними бібліографічними посиланнями з дотриманням норм академічної доброчесності.

**Мова та стиль викладення результатів.**

Рукопис дисертації підготовлено українською мовою на високому науково-технічному рівні. Виклад матеріалу відзначається чіткою логічною послідовністю та лаконічністю, що забезпечує цілісне сприйняття запропонованих інженерних і математичних рішень. Здобувач дотримується фахової термінології, що забезпечує точність і однозначність трактування викладених наукових положень.

Вступна частина дослідження сформована у повній відповідності до чинних нормативних вимог щодо структури кваліфікаційних наукових праць. У ній вичерпно обґрунтовано актуальність обраного напрямку, задекларовано зв'язок із науковими програмами та темами, чітко детерміновано мету, задачі, а також методи дослідження. Крім того, коректно структуровано пункти наукової новизни, практичної значущості роботи, відомостей про особистий внесок автора, стану апробації результатів та загальної характеристики публікацій.

**Скорочений опис розділів дисертації.**

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

розробки нових і 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 проєктування. 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 забезпечують 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* присвячений 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 для 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.

*У висновках* узагальнено головні результати дисертаційної роботи та окреслено напрями подальших досліджень щодо апаратного прискорення, оптимізації та реалізації ГЗМ на різнотипних мікрокомп'ютерних платформах

Дисертаційна робота оформлена відповідно до вимог наказу МОН України від 12 січня 2017 р. № 40 «Про затвердження вимог до оформлення дисертації».

**Оприлюднення результатів дисертаційної роботи.**

Загальні положення дослідження висвітлено в 17 наукових публікаціях, з них: 6 статтей в міжнародних періодичних фахових журналах, проіндексованих у базі даних Web of Science Core Collection та/або Scopus, з яких 2 статті у виданнях, віднесених до першого — третього квартилів (Q1—Q3) відповідно до класифікації SCImago Journal and Country Rank або Journal Citation Reports; 2 розділи в англомовних монографіях, 1 колективна монографія; 2 статті у наукових фахових виданнях України (які входять до переліку МОН України). Матеріали дисертаційної роботи доповідалися та обговорювалися на 6 міжнародних науково-технічних конференціях (IDAACS'2023, DESSERT'2023, ATIT'2020, AICT'2021, CMIS'2025 та ін.) та 3 Всеукраїнських наукових конференціях та семінарах,

Наукові публікації здобувача мають високий науковий рівень і пройшли відповідне рецензування. У кожну публікацію здобувач зробив вагомий особистий внесок, який був використаний при підготовці дисертаційної роботи. Усі публікації були написані з дотриманням принципів академічної доброчесності та складаються із власних досліджень здобувача та співавторів.

Таким чином, наукові результати описані в дисертаційній роботі повністю висвітлені у наукових публікаціях здобувача.

Опубліковані праці здобувача свідчать про його високий науковий рівень, що підтверджується актуальністю, оригінальністю та методичною глибиною досліджень. Вагомим показником якості є наявність 6публікацій, проіндексованих у наукометричній базі Scopus. При оприлюдненні результатів автор дотримувався норм академічної доброчесності: усі запозичені матеріали супроводжуються коректними посиланнями на першоджерела, ознаки плагіату відсутні. З огляду на це, можна стверджувати, що наукові положення та результати дисертації знайшли своє повне та вичерпне відображення в друкованих працях здобувача.

**Недоліки та зауваження до дисертаційної роботи.**

1. У першому розділі потрібно більше детально пояснити термін «кордонний пристрій», а також бажано деталізувати стратегію апаратно-параметричної оптимізації ГЗМ для мультимодального спектра застосування на кордонних пристроях в умовах їх апаратно-параметричних обмежень.
2. У яких одиницях оцінюється відстань Фреше між розподілами реальних та згенерованих зображень (на рис. 3.3 дисертації)?

3. У розділі 4 (Таблиця 3 на стор.113) **дає** порівняння алгоритмів для реалізації стратегії оптимізації трема показниками - Низкий, Середний та Высокий. Де кількісні оцінки?.
4. У п'ятому розділі при оцінці розгортання ГЗМ на мікрокомп'ютерній платформі Raspberry Pi 5 автор детально аналізує прискорення часу інференсу (до 0,61 мс) та зменшення обсягу моделі (до 0,55 МБ) завдяки INT8-квантуванню. Проте поза увагою залишилося питання оцінки енергоспоживання (у ватах або джоулях на один інференс) та температурного режиму процесора (троттлінгу) під час тривалого навантаження, що є критично важливими експлуатаційними параметрами для автономних кордонних пристроїв. Експериментальна верифікація апаратно-орієнтованої оптимізації ГЗМ здійснюється на базі мікрокомп'ютера Raspberry Pi 5. Хоча це рішення є цілком виправданим для класичних Edge-обчислень, дана платформа має порівняно потужну архітектуру ARM Cortex-A76. Зважаючи на стрімкий розвиток парадигми TinyML, доцільно було б доповнити дослідження аналізом можливості масштабування розроблених методів для інференсу ГЗМ на ультранизькоспоживаючих мікроконтролерах (наприклад, серії ARM Cortex-M із жорстким дефіцитом SRAM-пам'яті). Оскільки для таких платформ стандартного INT8-квантування зазвичай недостатньо, було б корисно дослідити, як запропонований автором гібридний фреймворк поєднується з методами екстремального апаратного стиснення (прунінгом нейромереж, дистиляцією знань або суббайтовим квантуванням INT4).
5. Процес динамічного квантування ваг моделі з формату FP32 у цілочисельний формат INT8 розглядається переважно як алгоритмічна процедура. З гляду на спеціальність «Комп'ютерна інженерія», доцільним було б додати формалізований опис відображення цих цілочисельних операцій безпосередньо на арифметико-логічні пристрої (ALU) мікропроцесорної архітектури ARM Cortex-A76, а також оцінити вплив промахів кеш-пам'яті (cache misses) на загальну затримку системи.
6. В тексті дисертації трапляються орфографічні та пунктуаційні помилки, (наприклад, на стор. 100 дисертації «… Сьогодні існує *безліч* модифікацій різних архітектур ГЗМ.»).

Підсумовуючи результати критичного аналізу рукопису, варто зазначити, що наведені зауваження переважно стосуються необхідності більш глибокої деталізації мікроархітектурних обчислювальних процесів або ж мають рекомендаційно-формальний характер. Інша частина коментарів

сформульована з метою розширення інженерного бачення здобувача щодо специфіки апаратно-програмного співпроєктування та окреслення перспективних шляхів подальшого масштабування розроблених систем.

Висловлені зауваження не є принциповими, жодним чином не нівелюють вагомості отриманих теоретичних і практичних здобутків, не знижують загального рівня наукової новизни та не впливають на підсумкову позитивну оцінку дисертаційної роботи.

**Висновок про дисертаційну роботу.**

Зважаючи на вищевикладене, вважаю, що дисертаційна робота здобувача ступеня доктора філософії Стрюка Олександра Сергійовича на тему «Оптимізація генеративних змагальних нейронних мереж в умовах апаратно-параметричних обмежень» є самостійним, цілком завершеним науково-технічним дослідженням. Робота виконана на високому науковому рівні із дотриманням принципів академічної доброчесності. Отримана сукупність теоретичних і експериментальних результатів успішно розв'язує актуальне науково-прикладне завдання, що має вагоме значення для розвитку комп'ютерної інженерії, зокрема у сфері проєктування вбудованих інтелектуальних систем.

За своєю актуальністю, рівнем наукової новизни та обґрунтованістю практичної цінності дисертація повною мірою задовольняє вимоги чинного законодавства України, 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:**
завідувач відділу мікропроцесорної
техніки № 205 Інституту кібернетики
імені В.М.Глушкова НАН України,
Лауреат Державної премії України,
доктор технічних наук, професор Володимир ОПАНАСЕНКО

Національна Академія наук України * Інститут кібернетики імені В.М.Глушкова * Ідентифікаційний код 05417176 * м.Київ * Україна *

04.05.2026

Підпис В. Опанасенко
ЗАСВІДЧУЮ
Зав. канц.
ІК НАН України