

Отже, інтелектуальні методи виявлення та реагування на кіберзагрози є ключовим напрямом розвитку сучасної кібербезпеки. Вони забезпечують швидкість, точність і масштабованість захисту, однак потребують подальшого вдосконалення, якісних даних та поєднання з експертною участю людини.

Список використаних джерел

1. Державна служба спеціального зв'язку та захисту інформації України. Офіційний вебсайт. URL: <https://cip.gov.ua> (дата звернення: 19.03.2026).
2. Національний координаційний центр кібербезпеки. Офіційний вебсайт. URL: <https://ncsc.gov.ua> (дата звернення: 19.03.2026).
3. CERT-UA. Офіційний вебсайт. URL: <https://cert.gov.ua> (дата звернення: 19.03.2026).
4. Закон України «Про основні засади забезпечення кібербезпеки України»: Закон України від 05.10.2017 № 2163-VIII. URL: <https://zakon.rada.gov.ua/laws/show/2163-19> (дата звернення: 19.03.2026).
5. Національний банк України. Офіційний вебсайт. URL: <https://bank.gov.ua> (дата звернення: 19.03.2026).

УДК 004.8:004.93

ЕФЕКТИВНІСТЬ QLORA ДОНАВЧАННЯ МУЛЬТИМОДАЛЬНИХ МОДЕЛЕЙ В УМОВАХ ОБМЕЖЕНОГО ОБЧИСЛЮВАЛЬНОГО БЮДЖЕТУ

*Цьома В. С.
tsyomav@gmail.com
Черкаський державний фаховий бізнес-коледж
Ночевнов Д.П.
м. Черкаси, Україна*

Мультимодальні мовні моделі набувають широкого застосування в задачах розуміння зображень, однак їх використання у спеціалізованих предметних областях потребує додаткової адаптації. Запуск моделей у повному розмірі вимагає значних обчислювальних ресурсів, тоді як менші моделі можуть не досягати достатньої точності у вузьких задачах класифікації. Донавчання

дозволяє адаптувати попередньо навчену модель до конкретної задачі без необхідності навчання з нуля. Проте вибір методу донавчання, архітектури моделі та гіперпараметрів залишається нетривіальною задачею, особливо в умовах обмеженого обчислювального бюджету.

Метою цієї роботи є порівняльний аналіз двох мультимодальних мовних моделей – Gemma 3 4B [1] та Qwen 3.5 4B [2] – у задачі класифікації зображень за допомогою методу QLoRA в умовах обмеженого обчислювального бюджету. Як тестова предметна область використовувалися зображення покемонів першого покоління – набір із 151 класу.

Для дослідження було зібрано набір даних із зображеннями покемонів. Початковий набір містив зображення на білому фоні, з приблизно 45 зображеннями на клас. З метою підвищення узагальнювальної здатності моделей та імітації реальних умов використання було проведено розширення даних: до кожного зображення застосовувалися випадкові комбінації накладання шуму, зміни перспективи, горизонтального відображення та незначного повороту. Кінцевий набір даних налічував близько 20000 зображень, що відповідає приблизно 135 зображенням на клас. Для оцінки якості моделей було окремо зібрано тестовий набір даних із приблизно 2300 зображень шляхом скрапінгу з пошукової системи Bing. Тестовий набір не перетинався з навчальними даними.

Для порівняльного аналізу було обрано дві мультимодальні мовні моделі – Gemma 3 та Qwen 3.5, обидві розмірністю 4 мільярди параметрів. Вибір обумовлений обмеженим обчислювальним бюджетом дослідження – одним GPU NVIDIA A100 40GB та 100 compute units у середовищі Google Colab Pro, що дозволило провести повний цикл навчання обох моделей у межах наявних ресурсів.

Повне донавчання, хоча й забезпечує максимальну адаптацію моделі, потребує значних обчислювальних ресурсів і тому не розглядалося як основний метод у цій роботі.

LoRA є параметрично ефективним методом донавчання, за якого тренуються лише додаткові низькорангові адаптери [3]. Подальшим розвитком

цього методу є QLoRA, що поєднує 4-бітне квантування базової моделі з навчанням LoRA-адаптерів. Це дає змогу істотно зменшити використання пам'яті без критичної втрати якості [4]. Саме тому поєднання цих двох методів було обрано для донавчання.

Для стабілізації навчання при рангу $r=32$ разом із QLoRA використовувався підхід rsLoRA [5], який покращує масштабування внеску адаптера, зменшуючи нестабільність навчання.

Перед навчанням набір даних було нормалізовано до формату ShareGPT, що використовується фреймворком Unsloth для мультимодального навчання, де кожен запис містить системний промпт, зображення та очікувану відповідь. Для обох моделей використовувалася однакова конфігурація QLoRA: базові ваги квантувалися до 4-бітної точності (NF4), адаптери LoRA застосовувалися до всіх лінійних шарів моделі. Основні параметри наведено в табл. 1.

Таблиця 1 – Значення заданих параметрів адаптерів

№	Параметр	Значення
1	Ранг (r)	32
2	Alpha (α)	64
3	Dropout	0.05
4	rsLoRA	Увімкнено

Навчання проводилося протягом 3 епох з однаковими гіперпараметрами для обох моделей.

На рис. 1 наведено криві втрат обох моделей протягом навчання у логарифмічному масштабі.

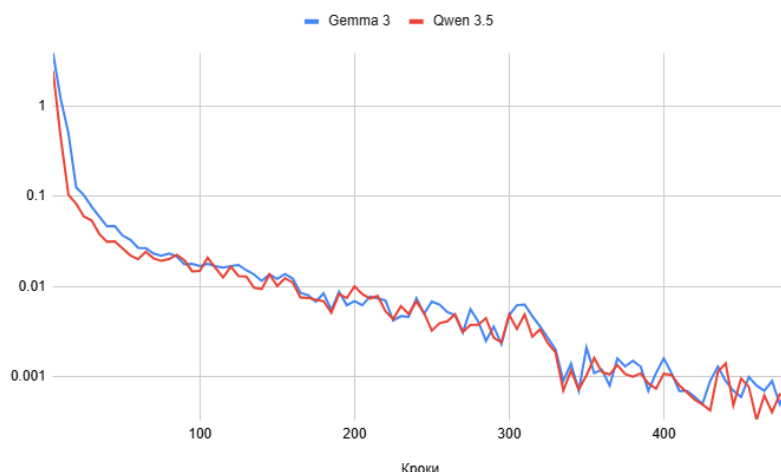


Рисунок 1 – Криві втрат Gemma 3 4B та Qwen 3.5 4B

Обидві моделі продемонстрували подібну динаміку зниження втрат і досягли близьких фінальних значень на навчальній вибірці. При цьому Qwen 3.5 показала дещо швидшу початкову адаптацію, що може свідчити про кращу придатність базової моделі до задач візуального розпізнавання.

Фінальним кроком стало тестування акуратності розпізнавання на попередньо підготованому датасеті. Результати наведено в табл. 2.

Таблиця 2 – Акуратність моделей на тестовому наборі даних

№	Модель	Епоха 1	Епоха 3
1	Gemma 3 4B	0.69%	82.2%
2	Qwen 3.5 4B	75.52%	80.92%

Отримані результати свідчать про швидшу початкову адаптацію Qwen 3.5 після першої епохи, тоді як Gemma 3, повільніше входячи в навчання, після трьох епох досягає вищої фінальної точності. Водночас Gemma 3 забезпечує швидший інференс, що робить її практичнішою за обмежених ресурсів.

Список використаних джерел:

1. Google Team. Gemma 3 technical report. 2025. URL: <https://arxiv.org/abs/2503.19786> (дата звернення: 22.03.2026).
2. Qwen Team. Qwen3 technical report. 2025. URL: <https://arxiv.org/abs/2505.09388> (дата звернення: 22.03.2026).
3. Hu E. J. та ін. LoRA: low-rank adaptation of large language models. 2021. URL: <https://arxiv.org/abs/2106.09685> (дата звернення: 22.03.2026).
4. Dettmers T. та ін. QLoRA: efficient finetuning of quantized LLMs // Advances in Neural Information Processing Systems. 2023. URL: <https://arxiv.org/abs/2305.14314> (дата звернення: 22.03.2026).
5. Kalajdzievski D. A rank stabilization scaling factor for fine-tuning with LoRA. 2023. URL: <https://arxiv.org/abs/2312.03732> (дата звернення: 22.03.2026).

УДК 004.056.3:004.65

КОМПЛЕКСНІ МЕХАНІЗМИ ЗАБЕЗПЕЧЕННЯ ВІДМОВСТІЙКОСТІ ДАНИХ У СИСТЕМАХ РЕЗЕРВНОГО КОПІЮВАННЯ

Павлишин А. І.

arturpavlishin6@gmail.com

Черкаський державний фаховий бізнес-коледж

Бреус Р.В.

м. Черкаси, Україна

У сучасних умовах стрімкого розвитку інформаційних технологій та зростання обсягів даних питання їх надійного зберігання набуває критичної важливості. Інформація є одним із ключових ресурсів організацій і користувачів, а її втрата може призвести до фінансових збитків, порушення бізнес-процесів і зниження рівня довіри. Основними причинами втрати даних є апаратні та програмні збої, кібератаки, помилки користувачів і зовнішні впливи [1–3].

У зв'язку з цим особливого значення набувають методи забезпечення надійності даних, серед яких базову роль відіграють резервне копіювання та відновлення. Комплексний підхід, що поєднує резервування, реплікацію,