

## ЕНЕРГОСПОЖИВАННЯ ТА ПРОДУКТИВНІСТЬ МОДЕЛЕЙ ШІ

Фесенко М.Б., Вітько О.В.

e-mail: mykyta.fesenko1@nure.ua, oleksandra.vitko@nure.ua

Харківський національний університет радіоелектроніки, каф. ШІ

м. Харків, Україна

The increasing computational complexity of modern artificial intelligence (AI) models has led to a significant rise in energy consumption. The training and deployment of large-scale neural networks require substantial computational power, raising concerns about environmental sustainability. This paper examines energy efficiency issues in deep learning, the role of optical processors in AI acceleration, and the comparative performance of GPU and TPU architectures in training neural networks.

Сучасні моделі глибокого навчання, зокрема великі мовні моделі (наприклад, GPT-4), вимагають значних обчислювальних ресурсів, що призводить до високого енергоспоживання та піднімає питання екологічної стійкості [1]. Традиційні кремнієві процесори, незважаючи на свою універсальність, стикаються з обмеженнями у продуктивності та енергоефективності. За деякими дослідженнями, навчання однієї великої моделі може спричинити викиди CO<sub>2</sub>, еквівалентні роботі тисяч автомобілів протягом року [1; 2].

Для зменшення енергоспоживання активно використовуються графічні процесори (GPU) та тензорні процесори (TPU). TPU, спеціально оптимізовані для виконання тензорних операцій, демонструють переваги у швидкості та зниженні споживання енергії порівняно з традиційними GPU. Проте, сучасні виклики стимулюють пошук нових рішень. Одним із перспективних напрямків є використання оптичних процесорів, які завдяки застосуванню фотоніки замість електроніки здатні суттєво знизити енергетичні витрати, зберігаючи високу швидкість обчислень, що є критичним для глибокого навчання.

Для оцінки співвідношення продуктивності моделі та її енергоспоживання було запропоновано показник Energy Productivity Index (EPI):

$$EPI = \lambda \cdot \ln \ln (P + 1) - \mu \cdot \ln \ln (E + 1)$$

де  $P$  – продуктивність моделі,  $E$  – споживання енергії,  $\lambda$  та  $\mu$  – вагові коефіцієнти. Підвищення значення  $P$  при одночасному зниженні  $E$  сприяє зростанню  $EPI$ , що свідчить про більш ефективне використання обчислювальних ресурсів.

Ключовим завданням сучасних досліджень є розробка енергоефективних алгоритмів, здатних оптимізувати процес навчання без втрати точності. Досягнення в галузі енергозбереження мають важливе

значення не лише з точки зору зниження експлуатаційних витрат, а й для мінімізації впливу на довкілля. Наприклад, оптимізація архітектур нейронних мереж, використання алгоритмів квантування параметрів та адаптивних методів розподілу навантаження допомагають знизити енергоспоживання при збереженні високої якості моделі [1; 3].

Сучасні технології, такі як оптичні процесори, відкривають нові перспективи у підвищенні енергоефективності [2]. Оптичні нейронні мережі здатні виконувати матричні операції з меншою витратою енергії, що дозволяє скоротити вуглецевий слід великих обчислювальних центрів. Порівняльний аналіз показує, що інтеграція таких рішень в архітектуру ШІ може забезпечити багатократне підвищення енергоефективності, що особливо актуально в умовах зростаючих обсягів даних і складності моделей.

Крім апаратних рішень, важливим є розвиток енергоефективних алгоритмів. Впровадження методів динамічного регулювання обчислювальних ресурсів, оптимізація процесів навчання та використання розподілених обчислень дозволяють мінімізувати витрати енергії без втрати продуктивності. Це сприяє не лише економії ресурсів, а й забезпечує стабільну роботу систем у реальних умовах експлуатації.

Подальші дослідження в цій галузі також охоплюють впровадження принципів «зелених» обчислень, коли енергія для дата-центрів надходить із відновлюваних джерел. Поєднання таких підходів з ефективною оптимізацією нейронних архітектур може ще більше зменшити рівень шкідливих викидів. Важливо, що підвищена увага до енергоефективності не лише стимулює технічний прогрес, але й формує суспільний запит на відповідальні технології, здатні гармонійно поєднувати високу продуктивність із турботою про довкілля.

Отже, інтеграція сучасних технологій – від TPU до оптичних процесорів, а також розробка нових алгоритмів оптимізації – є критично важливою для зниження енергоспоживання та підвищення продуктивності моделей ШІ [2; 3]. Баланс між продуктивністю та екологічною безпекою стане запорукою сталого розвитку штучного інтелекту у майбутньому.

#### Список використаних джерел:

1. Patterson D., Gonzalez J., Le Q., Liang C., Munguia L., Rothchild D. Carbon emissions and large neural network training. 2021. URL: <https://arxiv.org/abs/2104.10350> (дата звернення: 05.03.2025).
2. Shen Y., Harris N., Skirlo S. Deep learning with coherent nanophotonic circuits. *Nature Photonics*. 2017. Vol. 11. P. 441–446. URL: <https://doi.org/10.1038/nphoton.2017.93> (дата звернення: 05.03.2025).
3. Jouppi N., Young C., Patil N. In-Datacenter Performance Analysis of a Tensor Processing Unit. 2017. URL: <https://arxiv.org/abs/1704.04760> (дата звернення: 05.03.2025).