

Nvidia представила новые чипы серии Blackwell, которые значительно ускоряют процесс обучения больших моделей искусственного интеллекта (ИИ). Согласно свежим данным, опубликованным некоммерческой организацией MLCommons, количество необходимых чипов для тренировки крупных языковых моделей заметно сократилось.

MLCommons специализируется на сравнении производительности оборудования для искусственного интеллекта. В последний отчет были включены результаты тестирования чипов Nvidia и AMD, а также других производителей. При этом важным показателем остается эффективность именно обучения ИИ-моделей — этапа, когда системы получают огромные объемы данных для обучения. Хотя сейчас многие обращают внимание на скорость работы ИИ при ответах пользователям, именно количество и мощность чипов, задействованных в обучении, остается критичным фактором.

Тесты проводились на примере модели Llama 3.1 405B* — сложного открытого ИИ-модуля с огромным числом параметров, разработанного компанией Meta*. Данные о тренировке этой модели предоставила только Nvidia и ее партнеры. Результаты показали, что новые Blackwell-чипы вдвое быстрее прошлой серии Hopper на один чип.

Для завершения теста понадобилось 2496 Blackwell-чипов, которые справились с задачей за 27 минут. Это лучше показателей более чем в три раза большего количества предыдущих чипов.

Эксперты отмечают изменение подходов в индустрии: вместо огромных однородных кластеров из десятков тысяч чипов все чаще используются более компактные подсистемы для отдельных задач обучения ИИ. Такой метод позволяет ускорять процесс и сокращать время тренировки моделей с триллионами параметров.

* принадлежит компании Meta, организация признана экстремистской, её деятельность запрещена на территории России