

DeepSeek R2 выйдет позже запланированного из-за проблем с ИИ-чипами и дефицита GPU

По данным Financial Times, релиз новой модели DeepSeek R2 откладывается из-за неудачных попыток обучить её на чипах HUAWEI Ascend и дефицита высокопроизводительных GPU.

Источники сообщают, что компания сталкивается с техническими проблемами: обучение пока ведётся на NVIDIA, а чипы HUAWEI планируется использовать только для инференса.

HUAWEI, в свою очередь, даже отправила команду инженеров в офис DeepSeek для помощи, однако ни одной успешной сессии обучения на Ascend провести не удалось.

После выпуска R1 в январе, власти Китая настоятельно рекомендовали DeepSeek перейти на отечественные процессоры. Параллельно регуляторы вызвали крупные IT-компании, включая Tencent, ByteDance и Baidu, чтобы те объяснили закупки чипов NVIDIA H20.

По информации Mydrivers, R2 требует гораздо большего объёма обучающих данных, чем R1, но их качество на внутреннем рынке ниже. Вместе с острой нехваткой топовых GPU, это сильно замедлило процесс.

В результате, точная дата релиза R2 пока не назначена, а уровень использования сервисов DeepSeek с начала года упал с 50% до 3%.