



Компания представила **BitNet b1.58 2B4T** — крупнейшую на сегодняшний день модель ИИ с однобитной архитектурой. Этот открытый (по лицензии MIT) алгоритм способен работать даже на процессорах Apple уровня M2 благодаря архитектуре **bitnet**, специально оптимизированной для маломощного оборудования.

В основе **bitnet** лежит технология квантования весов модели до трех значений (-1, 0, 1), что значительно сокращает потребление памяти и вычислительных ресурсов.

BitNet b1.58 2B4T

содержит

2 миллиарда параметров

и обучен на

4 триллионах токенов

(примерно 33 миллиона книг). По производительности он сопоставим с аналогами от

Meta (Llama 3.2 1B), Google (Gemma 3 1B) и Alibaba (Qwen 2.5 1.5B)

, а в тестах

GSM8K

и

PIQA

даже превосходит их.

В некоторых случаях эта модель работает **в два раза быстрее** и требует меньше памяти. Однако для максимальной производительности ей необходим фреймворк

bitnet.cpp

, разработанный Microsoft. На данный момент он поддерживает только ограниченный набор аппаратных платформ и

не совместим

с графическими процессорами. Несмотря на перспективность архитектуры

bitnet

, именно ограничения по совместимости мешают ее более широкому распространению.