

VK открыла доступ к собственной нейросети RuModernBERT, предназначенной для обработки естественного русского языка. Эта модель может работать с длинными и сложными текстами без их деления на части, что делает её особенно полезной в задачах поиска, анализа и извлечения информации.

В отличие от аналогичных решений, RuModernBERT функционирует локально и не требует обращения к внешним API. Это снижает нагрузку на серверы и делает её более удобной для внедрения в различные продукты и приложения. VK уже интегрировала модель в свои сервисы, которыми ежедневно пользуются миллионы пользователей.

Нейросеть обучалась на массивном объёме данных — 2 триллиона токенов — на русском, английском языках и программном коде. При этом использовались разнообразные источники: книги, статьи, посты и комментарии из социальных сетей. Такая выборка позволила добиться понимания как формального, так и разговорного языка.

Модель доступна в нескольких вариантах: полная версия на 150 миллионов параметров и облегчённая на 35 миллионов. Это даёт возможность разработчикам выбирать решение в зависимости от конкретных задач и ресурсов. Кроме того, обновлены две дополнительные версии — USER и USER2, которые помогают искать и группировать похожую информацию. Во второй версии реализована технология, позволяющая значительно сократить объём данных без ощутимой потери точности.

С точки зрения производительности, RuModernBERT обрабатывает длинные тексты в 2-3 раза быстрее, чем предыдущая версия ModernBERT*, и ускоряет развертывание и обучение на конечных устройствах на 10-20%. По результатам внутренней оценки модель показала лучшие показатели среди аналогов в своей категории.