

Китайский ИИ бросил вызов гегемонии OpenAI и Anthropic.

Китайский стартап DeepSeek представил DeepSeek Coder V2 — первую открытую модель для программирования, превзошедшую по производительности закрытые аналоги GPT-4 Turbo, Claude 3 Opus, Gemini 1.5 Pro и опередившую Llama 3-70B.

Базирующаяся на предыдущей модели DeepSeek-V2, новинка демонстрирует выдающиеся результаты в кодировании и математических задачах, поддерживает свыше 300 языков программирования, сохраняя высокие способности к общему рассуждению и языковому пониманию.

Основанная в прошлом году с миссией «разгадать тайну AGI с помощью любопытства», DeepSeek стала значимым игроком на китайском рынке ИИ, присоединившись к таким компаниям, как SenseTime, Megvii Technology и Baidu. Фактически, в течение года после своего запуска стартап уже открыл доступ к ряду моделей, включая семейство DeepSeek Coder. Первая версия DeepSeek Coder, с числом параметров до 33 миллиардов, поддерживала 86 языков программирования и контекстное окно на 16 тысяч токенов. Новая версия V2 расширяет поддержку до 338 языков и увеличивает контекстное окно до 128 тысяч, что позволяет справляться с более сложными задачами программирования.

На тестах MBPP+, HumanEval и Aider, предназначенных для оценки генерации кода и решения задач, DeepSeek Coder V2 набрал 76.2, 90.2 и 73.7 баллов соответственно, опередив большинство моделей, включая GPT-4 Turbo, Claude 3 Opus, Gemini 1.5 Pro и Llama-3 70B. Аналогичные результаты были получены на тестах MATH и GSM8K, оценивающих математические способности модели.

Единственной моделью, превзошедшей DeepSeek Coder V2 на нескольких бенчмарках, стала GPT-4o, получившая более высокие баллы на HumanEval, LiveCode Bench, MATH и GSM8K. По словам разработчиков, таких результатов удалось достичь благодаря архитектуре Mixture of Experts и дополнительному обучению базовой модели DeepSeek V2 на 6 триллионах токенов, включающих программный код из GitHub и данные из CommonCrawl. Такой подход позволяет модели с 16 и 236 миллиардами параметров активировать только 2.4 и 21 миллиарда «экспертных» параметров для выполнения задач, оптимизируя вычислительные ресурсы и прикладные нужды.

Помимо выдающихся результатов в задачах программирования и математики, DeepSeek Coder V2 демонстрирует достойные результаты в общем рассуждении и понимании языка. На тесте MMLU, оценивающем понимание языка в различных

задачах, модель набрала 79.2 балла, что значительно лучше других моделей, ориентированных на кодирование, и почти аналогично результату Llama-3 70B. GPT-4o и Claude 3 Opus лидируют в категории MMLU с баллами 88.7 и 88.6 соответственно, в то время как GPT-4 Turbo следует за ними.

На данный момент DeepSeek Coder V2 предлагается под лицензией MIT, которая позволяет использовать ее как в исследовательских, так и в коммерческих целях без ограничений. Пользователи могут загрузить версии модели объемом 16 и 236 миллиардов параметров в базовом варианте и с дополнительными инструкциями через Hugging Face. Кроме того, компания также предоставляет доступ к моделям через свою платформу по модели оплаты по факту использования. Для тех, кто хочет сначала протестировать возможности моделей, компания предлагает опцию взаимодействия с DeepSeek Coder V2 через чат-бота.