

За последние несколько лет крупнейшие технологические компании активно улучшали свои большие языковые модели (LLM), что привело к быстрому внедрению таких продуктов в повседневную жизнь. Однако такие модели требуют огромных вычислительных мощностей, что ведет к высокому потреблению электроэнергии и значительным затратам на их поддержку.

В ответ на эту проблему исследователи стали обращаться к меньшим моделям, называемым SLM, которые являются менее ресурсоемкими и могут работать даже на локальных устройствах. Одним из главных способов использования SLM является их узкая специализация: вместо того чтобы отвечать на любые вопросы, они решают задачи в узкой области, такой как математика. Новый инструмент от Microsoft не только решает задачи, но и обучен логически рассуждать для поиска решений.

Модель rStar-Math, как отмечают разработчики, использует метод Монте-Карло для имитации человеческого подхода к решению задач поэтапно. Это позволяет модели разбивать задачу на составляющие, чтобы найти решение. Интересно, что rStar-Math выводит свой процесс размышлений как на естественном языке, так и в виде кода на Python.

Команда утверждает, что их модель успешно прошла несколько тестов и показала хорошие результаты.