



Исследование: удаление знаков препинания в запросе снижает точность языковых моделей до 20%

Грамотно сформулированные запросы повышают качество работы чат-ботов и виртуальных ассистентов.

Ученые Института AIRI обнаружили, что знаки препинания и артикли — элементы языка, которые часто воспринимаются пользователями как малозначимые, — играют ключевую роль в обработке информации языковыми моделями.

Языковые модели воспринимают текст как последовательность символов — токенов, среди которых системе необходимо выделить наиболее важные для понимания. В ходе исследования ученые разработали метод, который позволяет определить, какие именно данные сохраняются в связке с конкретными токенами.

Так, в первом эксперименте команда проверила, насколько хорошо специально разработанная языковая модель способна восстанавливать исходный текст. Для этого систему обучили анализировать последовательность токенов и восстанавливать текст, учитывая в том числе знаки препинания, артикли и стоп-слова — элементы текста, которые сами по себе не несут большого смысла. Оказалось, что больше всего информации о контексте содержится именно в стоп-словах.

Затем специалисты провели аналогичные тесты на наборах задач MMLU и BABILong, которые оценивают способности языковых моделей. В эксперименте использовались тексты, из которых заранее убрали элементы, кажущиеся незначительными для логики повествования. Чтобы определить, какие части можно исключить без потери смысла, ученые поставили перед языковыми моделями, включая ChatGPT, задачу — решить, какие элементы текста, с точки зрения человека, можно удалить. Проверка подтвердила, что, если убрать из формулировки технического задания такие «незначительные» символы, качество работы модели падает.

Код для анализа работы языковых моделей доступен по [ссылке](#).

«Изучение принципов действия языковых моделей — одна из главных задач нашей научной группы. Мы работаем над этим уже два года. Глубокое понимание того, как модели принимают решения, поможет сделать их не только эффективнее, но и безопаснее. Сейчас мы стремимся улучшить управляемость рассуждений языковых моделей, чтобы упростить их обучение и повысить



качество», – резюмировал **Антон Разжигаев**, руководитель группы «Интерпретируемый ИИ» лаборатории FusionBrain Института AIRI.

Вопросы: pr@airi.net

Институт AIRI — автономная некоммерческая организация, занимающаяся фундаментальными и прикладными исследованиями в области искусственного интеллекта. На сегодняшний день более 200 научных сотрудников AIRI задействовано в исследовательских проектах Института для работы совместно с глобальным сообществом разработчиков, академическими и промышленными партнерами.

Сайт: <https://airi.net/>