

JBS Dev: О несовершенных данных и «последней миле» ИИ — от возможностей моделей к устойчивости затрат

Источник: AI & Big Data Expo

Автор: Joe Rose

Оригинал: <https://www.artificialintelligence-news.com/news/jbs-dev-on-imperfect-data-and-the-ai-last-mile-from-model-capability-to-cost-sustainability/>

LLM

автоматизация

генеративный ИИ

медицинские данные

управление данными

Джо Роуз, президент стратегического технологического провайдера JBS Dev, хочет развеять один из мифов, связанных с работой генеративных и агентных систем искусственного интеллекта (AI). «Существует распространенное заблуждение, что ваши данные должны быть идеальными, прежде чем вы начнете выполнять любые виды подобных рабочих нагрузок», — объясняет он.

Как отмечается в недавней статье в AI Fieldbook, вендоры и консультанты — что неудивительно — предлагают соответственно использовать огромные «озера данных» (data lakes) и многолетние программы трансформации данных. В результате руководители находятся в замешательстве. Реальность несколько иная. «Инструментарий для работы с низкокачественными данными еще никогда не был лучше, чем сейчас», — говорит Роуз. «Поразительно, насколько хорошо LLM (Large Language Model — большая языковая модель) может понять даже наполовину написанный промпт».

Это логично. Если у вас есть такой инструмент, стоит использовать его в своих интересах — при условии наличия соответствующих защитных механизмов (guardrails). Внутренняя непредсказуемость моделей означает необходимость обработки некорректных выходных данных, и именно здесь в игру вступает принцип «человек в контуре» (human in the loop). Для текстовых или категориальных данных предусмотрена определенная устойчивость. «Люди привыкли к подходу: „мы это построили, оно работает, и мы об этом забываем“, — говорит Роуз. — Но эти системы работают совсем иначе».

Приводя пример работы с несовершенными данными, Роуз рассказывает о клиенте из медицинского сектора, целью которого была миграция в другую систему сверки счетов. Записи были разнородными: некоторые представляли собой PDF-файлы, другие — изображения; иногда процедура указывалась под именем врача, имя врача — под именем пациента, и так далее. Генеративный ИИ смог выделить чистые данные с помощью простого промпта: от OCR (Optical Character Recognition — оптическое распознавание символов) для изображений до извлечения текста из PDF. Впоследствии использовались более агентные подходы, такие как сравнение записи клиента со страховым контрактом для проверки правильности тарификации.

«Вы начинаете накладывать различные сценарии использования друг на друга», — говорит Роуз. «Это не означает, что система будет делать всё безошибочно — вам по-прежнему нужен человек в контуре. Но ваша цель — сказать: „мы начали с 20% автоматизации, затем перешли к 40%, затем к 60, 80%“, и постепенно наращивать этот показатель».

В будущем Роуз ожидает, что дискуссии вокруг этих моделей сместятся в сторону стоимости и переносимости. «Я думаю, вы увидите переход от обсуждения радикальных скачков и возможностей моделей к вопросам вроде: „как сделать стоимость более устойчивой, чтобы нам не пришлось строить дата-центры с той же скоростью, с которой мы их сейчас строим?“», — говорит он.

«Последний этап — это вопрос о том, как заставить эти вещи работать на ноутбуке или телефоне вместо того, чтобы запускать их в дата-центре. Модели обучались на массиве данных — по сути, на каждой странице в интернете и прочем. Не возникнет ситуации, когда появится еще тонна данных, которые еще не были в них загружены, и это приведет к какому-то прорыву».

На выставке AI & Big Data Expo, в которой участвует JBS Dev, Роуз с нетерпением ждет дискуссий. Еще одно его спорное мнение заключается в том, чтобы советовать людям перестать покупать решения у SaaS-вендоров (Software as a Service — программное обеспечение как услуга), если они могут реализовать это самостоятельно. «Это не так сложно, как кажется», — говорит он. «Почти у всех есть присутствие в облаке, и именно с этого я бы начал, потому что облачные инструменты, особенно у „большой тройки“... обладают всем необходимым, чтобы начать внедрение агентных рабочих нагрузок уже завтра, без покупки новых лицензий на ПО и дополнительного обучения».

Как только это будет реализовано, JBS Dev готова сопровождать клиента на следующих этапах этого пути.

Смотрите полное интервью с Роузом ниже:

Изображение: Gerd Altmann из Pixabay

Перевод выполнен: 15.05.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.