По мере того как ИИ-агенты берут на себя больше задач, управление становится приоритетом

Источник: Deloitte

Оригинал: https://www.artificialintelligence-news.com/news/as-ai-agents-take-on-more-tasks-governance-becomes-a-priority/

автономные системы | регулирование | риски | управление ИИ

Системы ИИ начинают выходить за рамки простых ответов. Во многих организациях сейчас тестируются **ИИ-агенты** (AI agents), способные планировать задачи, принимать решения и выполнять действия с ограниченным участием человека. Вопрос больше не заключается только в том, дает ли модель правильный ответ. Вопрос в том, что происходит, когда этой модели разрешают действовать.

Автономным системам необходимы четкие границы. Им нужны правила, определяющие, к чему они могут иметь доступ, что им разрешено делать и как отслеживаются их действия. Без такого контроля даже хорошо обученные системы могут создавать проблемы, которые трудно обнаружить или исправить.

Одной из компаний, работающих над этой проблемой, является **Deloitte**. Фирма разрабатывает механизмы управления (governance frameworks) и консультационные подходы, чтобы помочь организациям управлять системами ИИ.

От инструментов к ИИ-агентам

Большинство используемых сегодня систем ИИ все еще зависят от запросов человека (prompts). Они генерируют текст, анализируют данные или делают прогнозы, но человек обычно решает, что будет дальше. **Агентный ИИ** (Agentic AI) меняет эту модель. Эти системы могут разбивать цель на этапы, выбирать действия и взаимодействовать с другими системами для выполнения задач.

Такая дополнительная независимость создает новые вызовы. Когда система действует самостоятельно, она может выбирать пути, которые не были полностью предсказаны, или использовать данные способами, которые не были предусмотрены.

Работа Deloitte сосредоточена на том, чтобы помочь организациям подготовиться к этим рискам. Вместо того чтобы рассматривать ИИ как отдельный инструмент, фирма анализирует, как он вписывается в бизнес-процессы, включая то, как принимаются решения и как данные проходят через системы.

Внедрение управления в жизненный цикл

Управление не должно добавляться после развертывания. Оно должно быть встроено в полный жизненный цикл системы ИИ.

Это начинается на этапе проектирования. Организациям необходимо определить, что разрешено делать системе и где находятся ее пределы. Это может включать установку правил использования данных и описание того, как система должна реагировать в неопределенных ситуациях.

Следующий этап — развертывание. На этом этапе управление фокусируется на доступе и контроле, включая то, кто может использовать систему и к чему она может подключаться. Как только система вводится в эксплуатацию, основной задачей становится мониторинг. Автономные системы могут меняться со временем по мере взаимодействия с новыми данными. Без регулярных проверок они могут отклониться от своего первоначального предназначения.

Роль прозрачности и подотчетности

По мере того как системы ИИ берут на себя больше ответственности, становится все труднее отследить, как принимаются решения. Это создает спрос на более высокую прозрачность. Работа Deloitte подчеркивает важность отслеживания того, как работают системы. Это включает в себя протоколирование действий и документирование решений. Эти записи помогают организациям определить, что произошло, если что-то пошло не так. Если автономная система предпринимает действие, должна быть ясность относительно того, кто несет за это ответственность.

Исследование Deloitte показывает, что внедрение ИИ-агентов происходит быстрее, чем внедрение средств контроля, необходимых для управления ими. Около 23% компаний уже используют их, и ожидается, что этот показатель достигнет 74% в течение двух лет. Только 21% сообщают о наличии надежных мер защиты для надзора за их поведением.

Контроль ИИ-агентов в режиме реального времени

Как только автономная система активируется, основное внимание переключается на то, как она ведет себя в реальных условиях. Статических правил не всегда достаточно, и за системами необходимо наблюдать непосредственно в процессе их работы.

Подход Deloitte включает мониторинг в режиме реального времени, что позволяет организациям отслеживать действия системы ИИ во время выполнения задач. Если система ведет себя неожиданным образом, команды могут быстро вмешаться. Это может включать приостановку определенных действий или изменение прав доступа. Контроль в реальном времени также помогает в обеспечении соответствия нормативным требованиям (compliance). В регулируемых отраслях компаниям необходимо демонстрировать, что системы следуют правилам и стандартам.

На практике такие меры контроля начинают появляться в операционной среде. Deloitte описывает сценарии, в которых системы ИИ отслеживают работу оборудования на различных объектах. Данные с датчиков могут сигнализировать о ранних признаках сбоя, что может запустить рабочие процессы технического обслуживания и обновить внутренние системы. Механизмы управления определяют, какие действия может предпринимать система, когда требуется одобрение человека и как фиксируются решения. Процесс проходит через несколько систем, но с точки зрения пользователя это выглядит как единое действие.

Вопросы управления станут частью дискуссий на выставке **AI & Big Data Expo North America 2026**, которая пройдет 18–19 мая в Санта-Кларе, Калифорния. Deloitte указана как бриллиантовый спонсор (Diamond Sponsor) мероприятия, что ставит ее в ряд фирм, участвующих в обсуждении того, как автономные системы развертываются и контролируются на практике.

Задача состоит не только в создании более умных систем, но и в обеспечении того, чтобы они вели себя так, чтобы организации могли понимать, управлять ими и доверять им с течением времени.
