

# Генеративный искусственный интеллект – выгоды и риски

Арсений Ворошилов

**Д**искуссии о генеративном искусственном интеллекте (Generative AI) идут уже не первый год. Как и любая новая технология, эта тоже обрела как ярых сторонников, так и непримиримых противников, коих – тех и других – меньшинство по сравнению с практиками, которые рассматривают искусственный интеллект не как друга или врага, а как инструмент, который может быть полезен для решения определенных задач.

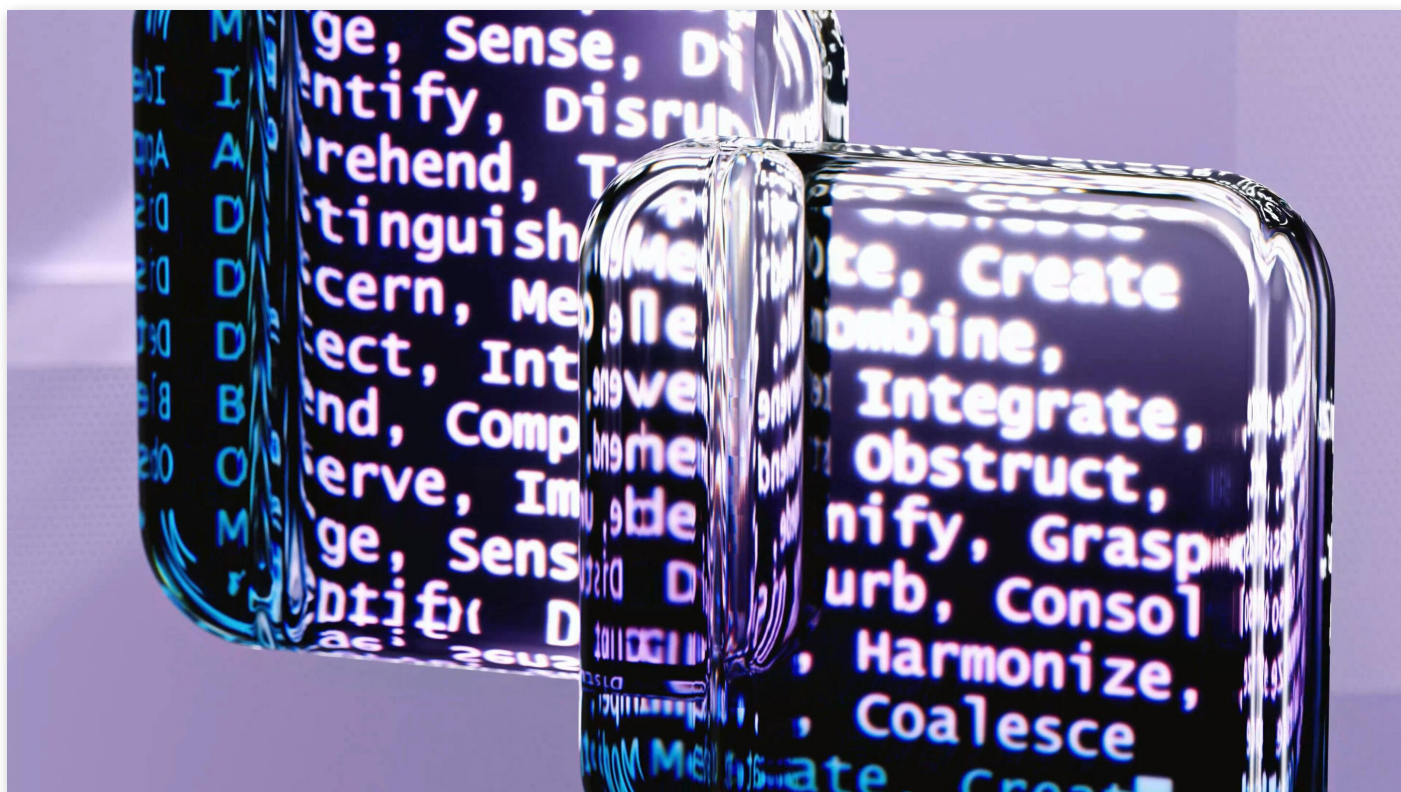
Но все это – применительно к уже готовым моделям, будь то Sora (недавно прикрытая ее создателем – компанией OpenAI), Google Gemini и любая другая. Однако, чтобы создать применимую на практике LLM-модель, ее надо сначала обучить, и здесь тоже применяется искусственный интеллект, причем не только для обучения, но даже для создания модели, которую предстоит обучать, ее тренировки, а также для выполнения определенных действий в системе машинного обучения.

Оказалось, что в использовании AI для решения всех этих задач кроются определенные риски. Такое мнение высказал ученый-компьютерщик Майкл Лоунз (Michael Lones). Хотя большие языковые модели LLM способны расширить возможности систем машинного обучения, сокращая одновременно расходы и трудозатраты, Лоунз предупреждает, что использование LLM чревато уменьшением прозрачности и эффективности управления для людей, разрабатывающих и использующих эти системы. Кроме того, возрастает опасность злонамеренных кибер-

атак, утечки данных и предвзятого отношения к недостаточно полно представленным группам.

*«Разработчики систем машинного обучения должны быть осведомлены о рисках, связанных с использованием GenAI в машинном обучении, и всегда находить точный баланс между улучшением возможностей и опасностями, которые этому сопутствуют, – говорит Майкл Лоунз, ученый в сфере компьютерных наук, работающий в Университете Хериота-Уатта в Эдинбурге (Великобритания). – С учетом нынешних ограничений, присущих генеративному AI, я бы сказал, что это яркий пример того, почему наличие возможности что-то сделать не означает, что это следует сделать».*

Имеет смысл разобраться в том, как происходит интеграция генеративного AI. Системы машинного обучения представляют собой алгоритмы, которые учатся распознавать закономерности (шаблоны) в данных, чтобы затем использовать их для предсказаний или принятия решений относительно новых данных. Машинное обучение существует уже несколько десятилетий, и большинство людей сталкивается с ним в ежедневной жизни. Это спам-фильтры, рекомендации товаров на интернет-сайтах электронной торговли, новостные ленты в соцсетях. В течение последних двух лет или около того наблюдалось движение в направлении встраивания генеративного искусственного интеллекта (в виде LLM) в системы машинного обучения, но такой подход несет риски и





ограничения, которые должны быть понятны и разработчикам, и обществу. Так считает Лоунз.

Он изучает четыре способа, с помощью которых генеративный AI в настоящее время применяется в машинном обучении: как компонент в составе конвейера машинного обучения; для разработки и кодирования конвейеров машинного обучения; для синтеза используемых в обучении данных; для анализа результатов машинного обучения. В каждом из этих случаев кроются риски, которые складываются, если LLM применяются для решения нескольких задач в системе машинного обучения, либо если LLM являются агентскими, то есть способны самостоятельно использовать внешние инструменты для решения той или иной задачи.

Еще одна проблема заключается в том, как полагает Майкл Лоунз, что если есть GenAI, который работает сразу по нескольким вариантам в рамках рабочего процесса или системы машинного обучения, то эти варианты (методы) могут взаимодействовать друг с другом непредсказуемо или малопонятно для человека. Поэтому Лоунз советует избегать избыточной сложности применительно к тому, как используется GenAI в машинном обучении, особенно если речь идет о секторе, функционирование которого существенно влияет на жизнь людей и на средства их существования. Несомненно, медиаиндустрия относится к таким секторам, поскольку человечество живет в информационную эпоху, а распространение генеративного искусственного интеллекта породило огромные объемы дезинформации, в том числе и аудиовизуальной, которая способна наносить большой вред обществу, подрывая его доверие к медиакомпаниям и иным информационным ресурсам.

К тому же один из наибольших рисков заключается просто в том, что LLM порой совершают ошибки, принимают неверные решения и фабрикуют так называемую «галлюцинаторную» информацию. Лоунз отмечает, что эти ошибки не обязательно предсказуемы и трудно под-

вергаются оценке, потому что LLM функционируют непрозрачно, что создает дополнительную сложность уже в смысле соблюдения законодательства.

*«В таких областях как медицина или финансы есть законы, обязывающие демонстрировать, что система машинного обучения надежна, и что пользователь способен объяснить, как она принимает решения, – говорит Лоунз. – Как только начинается применение LLM, это становится действительно трудно, потому что работа LLM предельно непрозрачна».*

Учитывая все сказанное, Лоунз советует разработчикам систем машинного обучения всегда вручную проверять сгенерированные LLM код и результаты. Он также предупреждает, что более крупные, дистанционно функционирующие LLM часто сохраняют данные и делятся ими, а значит, их использование создает возможности для нарушений кибербезопасности, утечки данных и иной чувствительной информации.

*«Для представителей широкой общественности важно знать об ограничениях, присущих системам GenAI, – говорит Лоунз. – Компании будут применять эти системы для достижения таких целей, как сокращение расходов, и это может принести определенную пользу конечным потребителям, но не исключены и негативные последствия, такие как предвзятость и несправедливость».*

Остается только еще раз повторить, что применительно к медиаиндустрии использование генеративного искусственного интеллекта как такового и в связи с машинным обучением несет те же риски, что и для других отраслей. Здесь тоже есть конфиденциальная информация, защищенные авторским правом данные, велик риск фейковых новостей, подтасовки и фальсификации важных фактов. Поэтому осведомленность о рисках позволяет более взвешенно применять алгоритмы искусственного интеллекта и лучше анализировать получаемые результаты.