ИИГЛА*

«Приносим свои извинения за неудобства, причиненные нашим клиентам. Мы считаем, что это недоразумение». Примерно так в Anthropic сообщили о первом в истории ИИ принудительном отзыве с рынка ИИ-сервиса по требованию правительства

* Как видно отечественная нейросеть не справилась с простейшим написанием заголовка «ИИгла». Что уж говорить, о доверии к иностранным сетям, которые могут не просто галлюцинировать, но и намеренно искажать результаты. К слову, американская нейросеть и вовсе отказалась генерировать иллюстрацию по заданному промпту, сославшись на свое несовершенство и правила.

Текст
ВАДИМ ФЕРЕНЕЦ,
ОБОЗРЕВАТЕЛЬ **«Б.О»**

«Чужая технология — это всегда чужой выключатель. Единственная гарантия, что завтра вас не оставят без ключевого инструмента, — собственный код, собственное железо и инженеры, которые понимают каждый его слой. Все остальное — аренда, которую могут отозвать в любой момент», — такое сообщение можно найти в телеграм-канале Консорциума исследований безопасности ИИ, созданного в 2024 году под эгидой Минцифры.

Это мнение экспертного сообщества появилось в связи с тем, что 12 июня 2026 года Правительство США, сославшись на закон о национальной безопасности и экспортный контроль, внезапно приказало компании Anthropic полностью отключить доступ к своим самым мощным моделям ИИ — Fable 5 и Mythos 5 — для всех пользователей по всему миру.

Концепция изменилась

Напомню, что IV Форум «Технологии доверенного ИИ», прошедший 13 мая 2026 года в Москве, в числе организаторов которого выступил упомянутый Консорциум, был в том числе посвящен тому, как выполнить поручение Президента РФ и внедрить ИИ во все ключевые отрасли, включая банковскую, к 2030 году, одновременно минимизировав риски.

По итогам дискуссии «Доверенный безопасный ИИ» в материале «Б.О» «Пакт о доверИИ» (по отношению к государству, людям и бизнесу) были описаны предложенные тогда экспертами три минимальных требования к ИИ для того, чтобы его можно было считать доверенным:

- ***техническая защита информации*** — классическая ИБ, защищенность от уязвимостей, закладываемых возможностей, ошибок и кибератак;
- ***функциональная надежность*** — способность системы ИИ стабильно, корректно и предсказуемо выполнять свои функции, выдавать правильные и воспроизводимые результаты;
- ***объяснимость*** — возможность понять и объяснить, как и почему ИИ принимает те или иные решения, что особенно важно для сложных моделей с большим объемом входных и выходных данных. Этот компонент тесно связан с вопросами этики и ответственности.

Прошло без малого два месяца, и ситуация, судя по всему, радикально изменилась. Какой смысл обеспечивать ИБ зарубежных сервисов, которые стали попросту недоступны или вот-вот таковыми станут? А вот для разработчиков отечественных моделей, совершенно очевидно, изменились приоритеты, о чем Консорциум заявил в публикации у себя на сайте.

Суровая реальность

«Кейс Anthropic — это не технический сбой, а инструмент политического принуждения, последствия которого мы обязаны просчитать», — утверждают авторы публикации и расставляют новые приоритеты.

1. ***Риск мгновенной технологической блокады.*** Как только ваш бизнес или государственная структура подсаживается на «технологическую иглу» иностранного сервиса, правообладатель получает «красную кнопку ИБ». Ее могут нажать под предлогом экспортного контроля, санкций или надуманной угрозы национальной безопасности США. Компания Anthropic была вынуждена подчиниться мгновенно, несмотря на несогласие, и это разрушило работу клиентов по всему миру.
2. ***Остановка критической инфраструктуры.*** Представьте, что подобная блокировка коснется не просто чат-бота, а промышленного ПО, облачных сервисов или систем управления базами данных (как это уже случалось с продуктами Oracle, Microsoft и SAP). Это грозит коллапсом логистики, финансового сектора и остановкой производств.
3. ***Непредсказуемость правил игры.*** Решения принимаются кулуарно. Anthropic прямо заявляет: им не предоставили детальных доказательств угрозы, процесс не был ни прозрачным, ни справедливым, ни основанным на технических фактах. Это означает, что у бизнеса нет возможности подготовиться, оценить риски или оспорить блокировку в правовом поле до того, как она произойдет.
4. ***Ловушка двойных стандартов.*** Пока технологический гигант тратил миллионы на пиар своей «безопасности», правительство США использовало именно эти заявления как основание для удара. Сэм Альтман, основатель и CEO компании OpenAI, которая является прямым конкурентом Anthropic на рынке ИИ, ранее назвал такую стратегию «маркетингом, основанным на страхе». Теперь же выяснилось, что тот, кто громче всех кричит о смертельной опасности своего оружия, первым попадает под регуляторный каток.

Вывод из ситуации с Anthropic однозначен: единственным условием технологического выживания является опора на открытые решения и национальные системы с собственными инженерными кадрами. Собственно, приказ ФСТЭК от 11 апреля 2025 года № 117 об этом прямо говорит. Именно по этой причине в России готовится ГОСТ, посвященный безопасной разработке моделей ИИ и т.д.

Ведь никто не сможет отключить пользователя от модели, если она развернута на его собственном оборудовании, работает на открытом исходном коде и обслуживается своими специалистами. ИИ перестает быть просто услугой — он становится элементом КИИ, которая должна быть под национальным контролем.

Рекомендации ЦБ

Одной из первых в мире на данный и иные кейсы, по информации IZ.ru, отреагировала Еврокомиссия, которая уже в начале июня 2026 года предложила проект закона Cloud and AI Development Act, предполагающего ограничение доступа к госконтрактам для американских IT-компаний.

Банк России 16 июня 2026 года опубликовал документ 3-МР «Методические рекомендации по обеспечению ИБ при разработке и применении ИИ на финансовом рынке», дополняющий наработки Консорциума. Ключевые рекомендации можно сгруппировать по нескольким направлениям:

1) ***управление рисками ИБ.*** Организациям рекомендуется проводить оценку рисков, связанных с нарушением ИБ и операционной надежности систем ИИ;
2) ***разработка модели угроз и мер защиты.*** Рекомендуется создавать модель угроз безопасности информации системы ИИ с учетом специфики технологий ИИ;
3) ***политика обеспечения ИБ.*** Организации следует разработать или дополнить внутреннюю политику по ИБ в сфере ИИ;
4) ***безопасность на этапах жизненного цикла ИИ.*** К ним относятся подготовка данных, разработка модели, обучение и тестирование, а также обеспечение надежности функционирования: шифрование входных и выходных данных, контроль аномалий, регистрация событий, ограничение параметров запросов, мониторинг штатной работы модели;
5) ***работа с поставщиками и открытыми компонентами.*** При использовании сервисов ИИ поставщиков или компонентов с открытым исходным кодом рекомендуется оценивать доверие к ним, проверять наличие у поставщика процедур аудита ИБ, требовать отчеты об анализе уязвимостей и включать в договоры положения об ответственности за инциденты;
6) ***реагирование на инциденты***. Необходимо регистрировать инциденты защиты информации и реагировать на них, а также иметь план восстановления операционной надежности на случай нештатных ситуаций.

Но если бы все ограничивалось только безопасностью ИИ! Сейчас ИБ становится заложницей разрушения и сужения глобального «пространства доверия», технологической базы обеспечения кибербезопасности. Именно здесь кроются причины хакерской атаки, совершенной 9 июня 2026 года против ГК «Астрал», включая удостоверяющий центр «Калуга-Астрал».

Кроме того, по словам известного ИБ-эксперта **Алексея Лукацкого,** массовый отзыв TLS-сертификатов 13 июня международным удостоверяющим центром GlobalSign есть не что иное, как публичное напоминание того, что HTTPS давно перестал быть чисто технической функцией веб-сервера. Публичный сертификат — это зависимость от глобальной инфраструктуры доверия, а эта инфраструктура живет по правилам юрисдикций, браузерных программ, аудиторов, санкционных офисов и риск-комитетов самих удостоверяющих центров. Теперь слово за Минцифрой. Б.О