1. С чего начать внедрение ИИ и что важно предусмотреть перед внедрением?

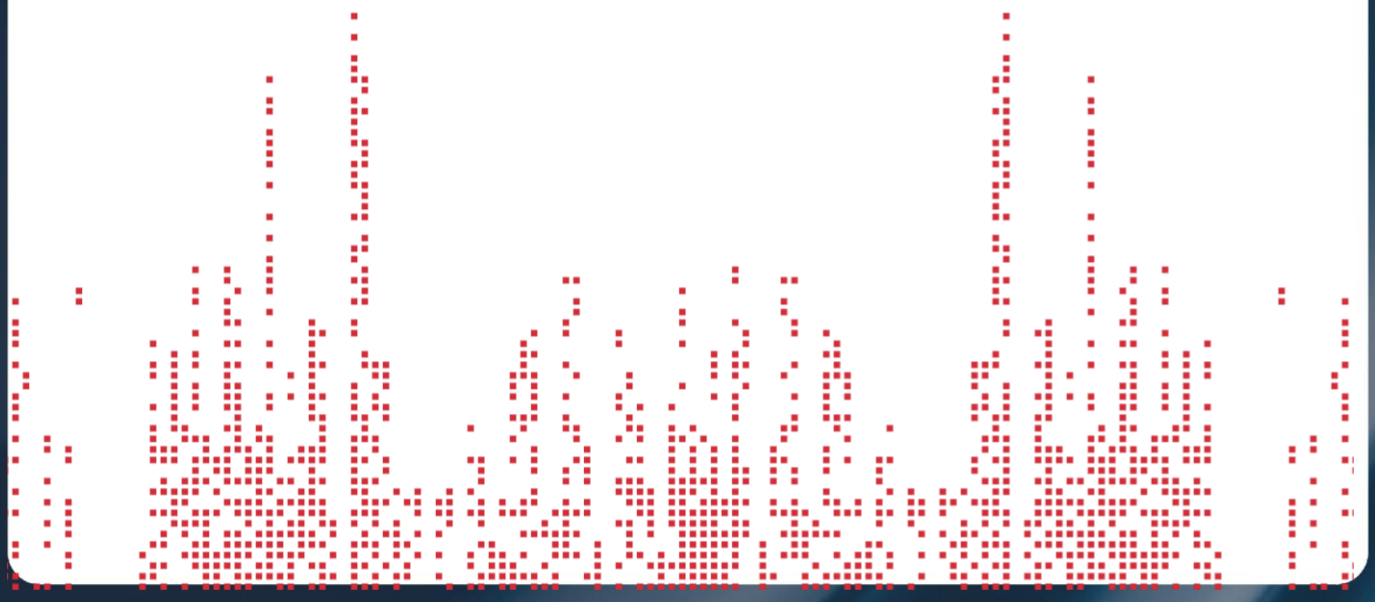
Внедрение ИИ – это прежде всего управленческий и юридический проект и лишь затем технологический.

Первый блок – внутренний аудит и ролевая самоидентификация. Компания должна определить, кем она выступает в каждом сценарии: разработчиком модели (проектирует, обучает или дообучает модель), оператором системы (эксплуатирует и настраивает систему, обрабатывает данные), лицом, предоставляющим доступ к сервису (обладает правами на сервис и открывает пользователям доступ через интерфейс), либо пользователем. Роли часто совмещаются, но чем сложнее архитектура (базовая модель одного поставщика, обёртка другого, интеграция третьего), тем важнее их разграничить: от роли зависят объём прав, состав обязанностей и распределение ответственности.

Второй блок – персональные данные и конфиденциальность. Загрузка в публичные модели (ChatGPT и аналоги) документов, содержащих персональные данные, влечёт риск нарушения Федерального закона от 27 июля 2006 г. № 152-ФЗ «О персональных данных»: как правило, отсутствует законное основание обработки, а сама передача данных зарубежному сервису квалифицируется как трансграничная и требует соблюдения ст. 12 указанного закона. До начала работы необходимо принять локальные нормативные акты, регулирующие порядок работы с данными, загружаемыми в нейросеть, а при поручении обработки вендору – оформить согласия и договор.

Третий блок – интеллектуальная собственность. Здесь два самостоятельных риска. На «входе» – использование при обучении и эксплуатации модели чужих охраняемых объектов: исключительные права на них сохраняются за правообладателями, а их использование без законного основания способно повлечь нарушение. На «выходе» – неопределённость правового режима самого сгенерированного результата: его охраноспособность и принадлежность прав на сегодня прямо не урегулированы, что требует договорного распределения прав на результаты генераций.

Четвертый блок – договорная архитектура и внутренние регламенты. Договор с вендором должен чётко распределять роли, фиксировать SLA, порядок обновлений и отката версий, условия обработки персональных данных, права на результаты генераций, ответственность, право на регрессное требование и обязанность содействовать при инцидентах. Внутри компании формируется комплект регламентов в зависимости от её роли – разработчика модели, оператора системы или пользователя сервиса.



2. С какими рисками сталкиваются компании, внедряющие ИИ?

Существует несколько ключевых групп рисков:

1. **Персональные данные и трансграничная передача.** Нейросети по умолчанию могут сохранять введённые персональные данные для обучения или передавать их третьим лицам. Обезличивание данных перед передачей в ИИ-систему не даёт надёжной защиты: совокупность, казалось бы, обезличенных сведений после их корреляции может вновь позволить идентифицировать субъекта. Риск усугубляется «теневой ИИ-экономикой»: сотрудники самостоятельно загружают конфиденциальные данные во внешние модели, а ответственность остаётся на компании.

2. **Дискриминация и нарушение прав человека.** Поскольку алгоритмы обучаются на исторических данных, они воспроизводят и усиливают нежелательные паттерны. Классический пример – экспериментальный рекрутинговый инструмент Amazon, отдававший предпочтение кандидатам-мужчинам, поскольку обучался на исторически мужской выборке. В трудовых отношениях подобные решения могут приводить к нарушению запрета дискриминации. Системы распознавания лиц позволяют отслеживать человека в публичных пространствах без его ведома и согласия.

Алгоритмы, анализирующие поведение пользователей в сети, формируют детальные психологические портреты, которые могут использоваться для таргетированного воздействия на решения и предпочтения человека. Алгоритмическая фильтрация контента незаметно управляет тем, какую информацию человек видит, а какую нет, ограничивая не только право на неприкосновенность частной жизни, но и свободу получения информации и свободу выражения мнений.

3. **Дезинформация, галлюцинации, дипфейки.** Генерация недостоверных сведений остаётся фундаментальным ограничением ИИ. Если компания принимает юридически значимое решение на основании галлюцинирующей модели, ответственность несёт она. Отдельную угрозу представляют дипфейки – синтезированные гиперреалистичные изображения и видео, способные вводить в заблуждение и использоваться для мошенничества и манипуляции.

4. **Интеллектуальная собственность.** При генерации ИИ способен нарушить чужие авторские права. Обучение модели на охраняемых объектах без разрешения правообладателей порождает риск претензий. В практике обсуждаются в том числе такие меры, как запрет использования и удаление неправомерно использованных данных.

5. **Кибербезопасность систем ИИ.** Уязвимость ИИ-систем к кибератакам образует самостоятельную группу рисков. Целенаправленные внешние атаки на нейросеть способны вызвать сбои в её поведении без ведома разработчика и пользователя: модель продолжает работать в штатном режиме, не «подозревая» об ошибке, а обнаружить такую атаку затруднительно. Одновременно и сам ИИ может содействовать вредоносной киберактивности – создавать фишинговый контент, автоматизировать социальную инженерию, выявлять уязвимости в инфраструктуре. В обоих случаях юридическая проблема общая: когда вред уже причинён, установить виновного и доказать причинно-следственную связь бывает сложно.

3. Кто отвечает за ошибку ИИ — разработчик, оператор, владелец сервиса или пользователь?

В действующем правопорядке ИИ не признаётся субъектом права и самостоятельной ответственности не несёт. Отвечает человек или организация, использовавшие технологию в своей деятельности. Объём ответственности участника ИИ-цепочки определяется его ролью, степенью реального контроля над ИИ-продуктом и способностью влиять на риск.

Статья 1064 ГК РФ закрепляет принцип генерального деликта, согласно которому вред подлежит возмещению в полном объёме лицом, причинившим вред. Для привлечения лица к ответственности в рамках генерального деликта в каждом случае необходимо доказывание состава гражданского правонарушения, включающего в себя четыре элемента:

1. ущерб (вред) и его размер с разумной степенью достоверности;
2. противоправность поведения причинителя вреда;
3. причинно-следственную связь между противоправным поведением причинителя вреда и наступлением вреда;
4. вину причинителя вреда.

Применительно к ИИ наибольшие затруднения вызывают именно последние два элемента: цепочка от действия человека до вредоносного результата может включать несколько автономных звеньев – алгоритм разработчика, самообучение модели, запрос пользователя, внешнее воздействие третьих лиц.

В результате ни один из участников не выступает очевидным единственным причинителем вреда, и суду приходится устанавливать, чьё именно поведение было противоправным и находилось ли оно в причинно-следственной связи с наступившим вредом.

Таким образом, формируется двухуровневая модель ответственности: внешний контур – отношения с клиентом, и внутренний контур – отношения между разработчиком, оператором и лицом, предоставляющим сервис.

Распределение ответственности внутри цепочки «разработчик – оператор – лицо, предоставляющее сервис» имеет ряд особенностей. Разработчик может отвечать не во всех случаях причинения вреда: ИИ способен к самообучению и постепенно отдаляется от первоначально заложенных алгоритмов; результат во многом зависит от формулировки запроса пользователя; возможно внешнее вредоносное воздействие – атаки на модель, при которых третьи лица целенаправленно искажают работу нейросети; наконец, исключительные права на модель могут принадлежать не разработчику. Полагаем, что разработчик должен отвечать перед потерпевшим лишь при одновременном соблюдении условий: он изначально заложил в модель свойство (дефект), приводящее к причинению вреда, предоставил возможность её использования иным лицам, и между базовым алгоритмом и причинившим вред результатом существует прямая причинно-следственная связь.

Например, перед пациентом отвечает клиника по законодательству об основах охраны здоровья граждан, а при оказании платных услуг также по законодательству о защите прав потребителей. Однако внутри цепочки клиника (как оператор, а нередко и лицо, предоставляющее сервис) вправе предъявить регрессное требование разработчику модели, если докажет, что ошибка изначально была заложена в свойствах самой модели и причинно-следственную связь с наступлением вреда. Напротив, если клиника проигнорировала предупреждения разработчика об ограничениях модели либо не внедрила процедуру верификации результата человеком, распределение вины сместится в её сторону.

Особый режим установлен для государственных органов: вред, причинённый при использовании ИИ в деятельности правоохранительных и иных госорганов, возмещается за счёт казны по правилам ст. 1069–1070 ГК РФ независимо от ведомственной принадлежности технологии. Данные нормы не учитывают специфику ИИ и, вероятно, потребуют развития, однако сам принцип ответственности государства как конечного пользователя технологии применим уже сегодня.

4. А как складывается судебная практика в отношении ответственности за действия ИИ в России и за рубежом?

В сложившейся практике есть несколько примечательных кейсов, в которых перед судом вставал вопрос об ответственности субъектов за действия ИИ.

Китай: Li Xiaoliang v. Baidu – клевета, сгенерированная ИИ

Интересно дело Li Xiaoliang v. Baidu, рассмотренное Нанкинским народным судом промежуточной инстанции Китая. Адвокат из Нанкина обнаружил, что при поисковом запросе его имени сервис Baidu AI выдавал ответ, из которого следовало, что он якобы был осужден к лишению свободы, и сопровождал результат его фотографией в адвокатской мантии. Компания Baidu возразила, заявив о невозможности контролировать работу алгоритма, и сослалась на раннюю стадию развития самой технологии. Суд отметил, что указанные ограничения не наблюдаются у других ИИ-платформ, что свидетельствует о принципиальной возможности контроля над алгоритмами. Суд разграничил обычные поисковые подсказки, и «умный ИИ-ответ», когда порочащий текст был соединен с изображением истца, – и в этом синтезе суд усмотрел вину Baidu. Суд обязал Baidu опубликовать решение суда, но при этом отказал во взыскании имущественных убытков и компенсации морального вреда. Данное дело уже назвали первым в Китае судебным спором о нарушении, совершённом большой ИИ-моделью.

Канада: Moffatt v. Air Canada – ответственность за чат-бот на сайте

В деле Moffatt v. Air Canada, 2024 BCCRT 149 чат-бот авиакомпании на сайте дал пассажиру неверную информацию о политике возврата билетов, пообещав скидку, которой не было. Суд обязал авиакомпанию выплатить компенсацию, указав, что компания несет полную ответственность за информацию на своем сайте независимо от того, кто ее сгенерировал – человек или ИИ. Интересно, что авиакомпания приводила в свою защиту довод о том, что ИИ является отдельным субъектом прав, но суд отверг этот аргумент, указав, что чат-бот выступает в качестве составляющей сайта.

США: Garcia v. Character Technologies – ответственность разработчика, оператора и провайдера инфраструктуры

В деле Garcia v. Character Technologies, Inc. (M.D. Fla., 2025) мать погибшего подростка предъявила иск к разработчику чат-бота, его основателям и Google. Суд при разрешении дела указал, что разработчик может нести строгую ответственность за дефекты дизайна ИИ-системы, причинившей реальный вред. Кроме того, в судебном решении указано, что у компании-оператора ИИ-сервиса есть обязанность заботы о пользователях, поскольку выпуск системы в открытый доступ создает предвидимые риски.

Суд отказался квалифицировать ответы бота как речь и распространять на нее защиту, предусмотренную первой поправкой к Конституции США, тем самым отклонив аргумент о том, что разработчик и оператор не несут ответственности за слова ИИ.

Наконец, суд допустил привлечение к ответственности Google как поставщика инфраструктуры. Таким образом, разработчик и оператор ИИ в лице Character Technologies, Inc. и провайдер мощностей в лице Google LLC были признаны надлежащими ответчиками. Ссылка на то, что правонарушение совершено ИИ, а не компанией, не освобождает использующих его юридических лиц от ответственности.

Россия: апелляционное определение Тверского областного суда – ответственность владельца интернет-ресурса

В российской практике формируется собственная практика. Так, Апелляционное определение Тверского областного суда от 5 декабря 2024 г. № 2-34/2024 (УИД 69RS0032-01-2023-002632-49) демонстрирует, что довод о том, что спорные публикации «были созданы искусственным интеллектом автоматически», не освободил владельца информационного портала от ответственности. Суд взыскал с него компенсацию морального вреда за оскорбительный характер материалов, сгенерированных ИИ, подчеркнув, что размещение такого контента на принадлежащем ответчику ресурсе влечет ответственность по общим правилам ГК РФ.

Буду рада ответить на ваши вопросы

Елена Полевая

Старший юрист

Elena.Polevaya@kkmp.legal