Регулирование на пересечении ИИ и ИБ

как часть управляемой системы

Максим МИЛКОВ,
руководитель направления «Искусственный интеллект» «Софтлайн Решения» (ГК Softline)

Искусственный интеллект за последние годы перестал быть «пилотной технологией», которую можно осторожно обкатывать в отдельных подразделениях, не меняя архитектуру управления рисками. Теперь как часть базового ИТ-стека ИИ встраивается в принятие решений, от поддержки операционных процессов до агентных сценариев, где система сама инициирует действия. На этом фоне нормативно-правовая база в области пересечения ИИ и информационной безопасности развивается не как отвлеченная «этика технологий», а как прагматичный ответ на рост масштабов и расширение последствий применения моделей. Регулятор видит не красивую витрину инноваций, а новую поверхность атаки, дополнительный класс критических активов и формы ущерба, которые невозможно закрывать только внутренними политиками компаний.

Если испортить данные на входе, система ошибется на выходе

Связка ИИ и ИБ выглядит естественной: методы искусственного интеллекта опираются на данные, а данные – это актив, который защищается государством и бизнесом и который все чаще признается стратегическим ресурсом. Если испортить данные на входе, система начнет ошибаться на выходе. Если атаковать модель, она может выдавать вредоносные рекомендации, утекать через интерфейсы, поддаваться манипуляциям. Если переносить решения, которые раньше принимались человеком, на автоматизированный контур, то цена ошибки, сбоя или злонамеренного влияния кратно возрастает.

Неслучайно регулирование в зоне пересечения ИИ и ИБ почти везде развивается через три взаимосвязанные сущности: правила работы с данными как с «сырьем» для моделей, требования к ИИ-системе как к объекту защиты и через контроль последствий решений, которые принимаются с участием ИИ или передаются ему полностью.

Если говорить о российском контуре, то в стране пока нет единого «зонтичного» закона, который комплексно охватывает искусственный интеллект и сопутствующие риски информационной безопасности. Действует эволюционный подход: государство усиливает и дополняет существующие массивы регулирования, которые исторически отвечали за данные, критические информационные системы и безопасность в госсекторе. Такой кажущийся консервативным путь отражает реальность: значительная часть рисков ИИ укладывается в привычные рамки ИБ, если правильно расширить понятия угроз и инструментов контроля.

Рост требований к технологическим мерам защиты

В сфере данных в России базовым и наиболее «нагруженным» остается регулирование обработки персональных данных. Этот старый и массивный контур регулярно пересматривается, через него проходят многие ИИ-проекты, потому что любой прикладной ИИ упирается в персональные или квазиперсональные атрибуты. Ключевой вектор последних лет – не декларативные требования «защищать и обезличивать», а попытка сделать обезличивание проверяемой технической дисциплиной, чтобы снизить обратимость преобразований и риск восстановления исходных персональных данных при утечке.

Показательным в этом смысле является приказ «Роскомнадзора» № 140. Его принципиальность не в том, что им «введено обезличивание», концепция существовала. Сдвиг в том, что регулятор

попытался зафиксировать единые методологические требования и приблизить обезличивание к набору конкретных инженерных подходов, которые можно описать, воспроизвести и проверить. По сути, если организация использует датасеты для аналитики или обучения моделей, она должна не просто заявлять про обезличивание, а подтверждать, какими способами снижается риск обратной идентификации. На практике это означает повышение требований к технологическим мерам защиты, журналированию операций, аудиту действий с данными, наличию методик и внутренних регламентов, которые можно предъявить проверяющей стороне. Для крупного бизнеса это уже не факультативная бюрократия, а отдельная статья затрат и процессов, причем независимо от того, позиционирует компания свой проект как ИИ или как классическую аналитику.

Второй важный российский сюжет – формирование государственного «озера данных» и институционализация доступа государства к обезличенным датасетам. Появление в 152-ФЗ требований, связанных с передачей обезличенных наборов в ГИС Минцифры, можно интерпретировать как следующий виток политики суверенитета данных. Суть механизма не в том, что компании обязаны «по умолчанию» выгружать все, чем располагают, а в том, что при запросе со стороны госоргана через Минцифры организация обязана предоставить обезличенные данные. В результате государство конструирует легальный контур накопления датасетов, который позволяет обогащать государственные массивы данными, собранными частным сектором, и использовать их для построения моделей и сервисов на уровне страны. Прослеживается общая линия: сначала закреплялась локализация персональных данных на российской инфраструктуре, затем усиливались контроль и ответственность за обработку, а теперь дополнительно формируется возможность государственно-централизованного доступа к обезличенным наборам.

На практике это означает, что компании должны обеспечить техническую безопасность обработки и быть готовыми к тому, что часть данных, пусть и в обезличенном виде, может стать элементом более широкой государственной инфраструктуры.

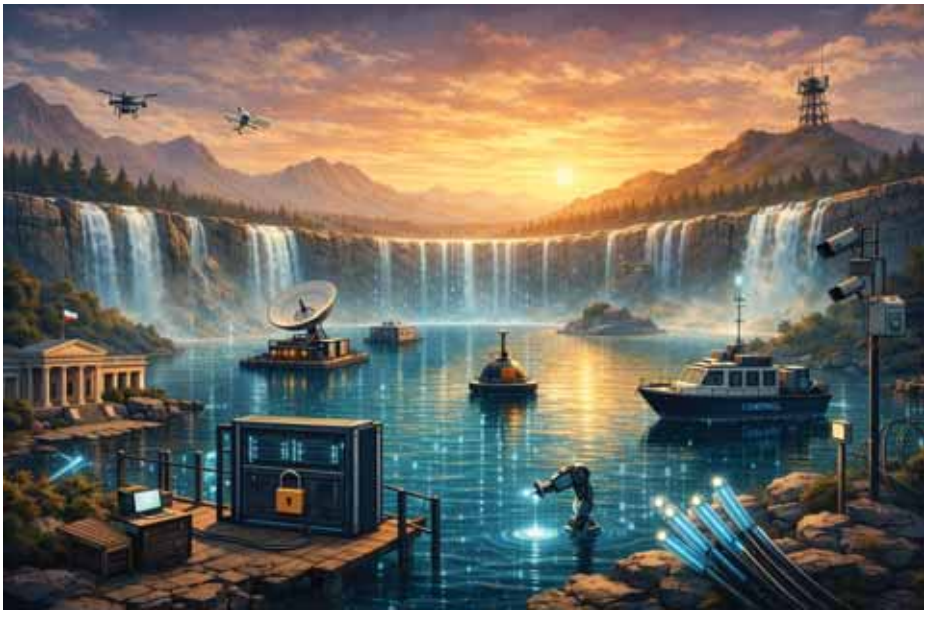
ИИ перестает восприниматься как еще один ИТ-продукт

В данных государство усиливает контроль, а в ИИ-системах фиксирует новый объект регулирования – модель и ее жизненный цикл. Здесь логика вокруг традиционного для ИБ вопроса: где максимальная критичность и где цена ошибки недопустима. Отсюда естественный фокус на госсекторе и на критической информационной инфраструктуре. В этой зоне ИИ перестает восприниматься как еще один ИТ-продукт. Он рассматривается как специфический инструмент со своей моделью угроз и типовыми уязвимостями, который должен проходить отдельные процедуры оценки, испытаний и допуска.

На этом фоне опорой выступает приказ ФСТЭК № 117, который закрепляет требования к системам в госсекторе и задает направления, куда будет развиваться нормативка. Особенно показательна идея запрета дообучения моделей «на лету». Регулятор исходит из того, что модели, которые постоянно изменяются в продуктиве, резко усложняют оценку рисков и дают слишком много пространства для манипуляций, деградации качества и появления неучтенных поведений. Поэтому предпочтение отдается «жестким» моделям: набор данных собран, обучение проведено, результаты валидированы, риски оценены, используется фиксированная версия.

Второй пласт – аттестация. Если система влияет на существенные процессы в госсекторе, то перед внедрением должны появляться аккредитованные процедуры проверки, аналогичные тем, что давно применяются к определенным классам программных и технических средств. ИИ встраивается в эту матрицу как новый объект сертифицируемой безопасности.

Дополнительным элементом служит контур мониторинга инцидентов, связанный с ГосСОПКА: инциденты должны фиксироваться, обрабатываться и передаваться по установленным правилам. В совокупности с требованиями по отказоустойчивости,



резервированию и управлению эксплуатацией это означает, что внедрение ИИ в госсекторе и критичных сегментах экономики превращается в проект с заметно большей долей инженерной и регуляторной работы, чем это характерно для «обычных» ИТ-систем. Интегратор и вендор должны планировать не только разработку и внедрение, но и доказательную базу безопасности, подготовку документации, прохождение оценок и готовность к проверкам.

Государство систематизирует требования к идентификации угроз

Российский контур развивается не только через приказы ведомств, но и через стандартизацию, которая пока находится в стадии формирования. Появление проекта ГОСТа, описывающего использование ИИ в классической инфраструктуре, – важный сигнал: государство систематизирует требования к идентификации угроз, качеству данных, интерпретации результатов моделей и тому, какие атаки способны негативно повлиять на поведение ИИ-систем.

На текущем этапе в таких документах преобладают общие механизмы защиты, похожие на классические ИТ-подходы: верификация, резервирование, экспертиза, аудит. Однако логика стандартизации почти всегда одинакова: сначала фиксируется верхнеуровневый «скелет» требований, который обрастает отраслевыми ответвлениями и детализацией. ИИ в этом смысле не исключение. Как только практика внедрения и инциденты дадут достаточно материала, требования станут более точными, особенно в части тестирования моделей и специфических сценариев атак, характерных для генеративных и агентных систем.

Таким образом, российская траектория выглядит как постепенное «приземление» ИИ на давно известные рельсы информационной безопасности с необходимым расширением понятий. Государство усиливает суверенитет данных, повышает контроль процедур обработки, формирует механизм доступа к обезличенным наборам, а в критичных и государственных сегментах начинает выделять ИИ как отдельный класс систем, для которых нужны свои профиль угроз, процедуры испытаний и доказательная база безопасности. Ближайшая перспектива здесь – не появление внезапного супертекста «закона об ИИ», а вступление в силу уже подготовленных требований и переход к контролю на практике: проверкам, аттестациям, аудитам. Это означает, что проекты, рассчитанные на 2026–2027 гг., должны закладывать дополнительные ресурсы на комплаенс и безопасность не в виде абстрактных мер, а в виде работ, документов, тестов и процедур.

Китайский механизм и идея технологического суверенитета

Зарубежный опыт не сводится к одному универсальному сценарию. Он показывает, что государства могут идти двумя принципиально разными маршрутами, оставаясь при этом в одной логике: контроль рисков там, где последствия критичны. Показательный пример – Китай, где, как и в России, нет единого глобального закона, «зонтично» регулирующего весь ИИ. Но в Поднебесной используется набор специализированных указов и требований, которые точечно накрывают отдельные классы ИИ-систем. В этом есть общие черты с российским опытом, например, внимание к локализации данных и контролю критичных сегментов, а фундаментальные отличия – в степени жесткости и роли государства.

Китайский механизм во многом строится вокруг идеи технологического суверенитета, где государство поддерживает развитие фундаментальных исследований, собственных моделей, алгоритмов и инфраструктуры. Одновременно вводится обязательная регистрация алгоритмов в сегментах, где их влияние может затрагивать общественно значимые процессы. На практике это касается рекомендательных систем и информационных лент, т. е. того, где алгоритм способен влиять на общественное мнение и масштабное поведение пользователей. Иными словами, государство стремится контролировать этот слой алгоритмического управления, рассматривая его как критический.

Выделяется требование к аудиту и оценке перед запуском: для некоторых классов систем предполагаются полноценная проверка и допуск, занимающие значимое время. Это неминуемо снижает скорость инноваций в регулируемых сегментах, но государство сознательно платит высокую цену ради управляемости. Важнейшая особенность китайской модели в том, что регуляторные ограничения часто носят не финансовый, а эксплуатационный характер: сервис может быть заблокирован. Если в российской логике нарушение чаще упирается в штрафы и предписания, то в китайской – в возможность остановки работы инструмента. Тем самым формируется иной баланс сил: риск для бизнеса не «заплатим и исправим», а «можем потерять доступ к рынку и каналам распространения».

Китай демонстрирует строгий подход к проверке обучающих выборок, в том числе на предмет предвзятостей, и к регулярной переаттестации систем. В регуляторной конструкции появляется механизм, при котором государство может требовать отключения модели или сервиса. Это укрепляет ощущение, что в отдельных сегментах ИИ рассматривается не просто как частный технологический инструмент, а как часть государственно контролируемой инфраструктуры. В то же время китайская система не копирует российскую в части некоторых прав субъектов данных. В российском регулировании есть норма,

позволяющая по желанию удалить персональные данные из набора, в китайской логике такая опция выражена существенно слабее. В части обезличивания российские акты выглядят более детализированными методологически, но китайский подход компенсирует это глубиной проверки в отдельных отраслях, где государство может фактически требовать предоставления выборок на оценку безопасности и допустимости.

Усиление тренда на суверенитет данных

Если сравнить российский и китайский подходы как управленческие модели, то Россия в текущей точке воспринимает ИИ прежде всего как новый класс систем, расширяющий поверхность атаки и требующий включения в привычные контуры киберустойчивости, аудита, логирования и сертификации, особенно в госсекторе и критической инфраструктуре. Китай, напротив, сильнее тяготеет к модели, где ИИ в критичных общественно значимых сегментах воспринимается как инструмент, который должен быть встроен в государственный контур контроля и допуска. Разница не в наличии или отсутствии «общего закона» – его нет, – а в степени концентрации полномочий и характере ограничений.

Перспективы в России на ближайшие годы выглядят предсказуемо, поэтому требуют внимания уже сейчас. Первое – это завершение оформления стандартов и ведомственных требований и их переход из статуса «анонса» и «проектов» в статус обязательной практики. Второе – появление более конкретных требований к тестированию моделей и доказательству безопасности, особенно для генеративных систем, которые обладают специфическими уязвимостями, плохо покрываемыми классическими чек-листами ИБ. Третье – институционализация аттестации ИИ-систем по аналогии с тем, как сертифицируются средства защиты и определенные классы программных продуктов, используемых в чувствительных контурах. У заказчиков и исполнителей появятся новые обязательные этапы жизненного цикла, которые нельзя будет «срезать», не утратив возможности внедрения.

Одновременно будет усиливаться тренд на суверенитет данных, который проявился через локализацию и механизм доступа государства к обезличенным наборам. На этом фоне вероятно, что крупные компании начнут активнее развивать практику синтетических данных как компромиссный инструмент. Синтетика позволяет обучать и валидировать модели на данных, которые статистически отражают реальные закономерности, но не несут прямой ценности в виде исходных персональных записей. В условиях, когда по запросу государства компании обязаны передавать обезличенные датасеты, синтетические наборы могут стать способом одновременно поддерживать качество моделей и снижать регуляторные и репутационные риски. Это не отменяет требований к безопасности, но меняет структуру риска: компания меньше зависит от «сырья» в виде первичных персональных массивов и, следовательно, снижает чувствительность проекта к утечкам и спорным ситуациям вокруг доступа к данным.

Глобальные тренды нормативного развития в зоне ИИ и ИБ сводятся к трем полям борьбы. Первое поле – данные: их качество, происхождение, способы обезличивания, безопасность хранения и юридический режим доступа. Второе – киберустойчивость ИИ-систем: способность противостоять новым типам атак и иметь понятные процедуры тестирования, обновления и контроля версий. Третье – прозрачность процессов использования: логирование, журналирование, аудит, доказуемость принятия решений и возможность восстановления цепочки событий в случае инцидента. Остальное, включая дискуссии о гуманитарных аспектах ИИ, в прикладном контуре информационной безопасности превращается в требования к процессам и к доказательной базе.

Для рынка это означает, что зрелость ИИ-проектов будет измеряться не только точностью моделей и экономическим эффектом, но и тем, насколько проект подготовлен к требованиям регулятора и реальным инцидентам. В сегментах госсектора и критической инфраструктуры «быстро внедрить» будет все труднее без инженерной дисциплины и заранее выстроенной нормативной рамки. Для коммерческих компаний вне КИИ ключевым риском становится не столько прямой запрет, сколько повышение ответственности за данные и расширение полномочий проверяющих органов, что будет заставлять бизнес инвестировать в процессы обработки данных, документацию и аудит так же системно, как и в инфраструктуру.

ИИ становится «обычной» технологией

Таким образом, разговор о нормативно-правовой базе пересечения ИИ и ИБ – это не про будущее, а про настоящее. ИИ становится «обычной» технологией, как когда-то стали обычными «облака» и мобильные платформы. Но в отличие от прежних волн цифровизации у него другая природа риска: он влияет на решения, меняет управляемость процессов, создает новую поверхность атаки, а значит, неизбежно становится объектом регулирования.

Российская модель делает ставку на эволюцию существующих норм и на фокус в госсекторе и критичных отраслях, Китай – на точечный, но гораздо более жесткий контроль в социально значимых сегментах. Обе траектории демонстрируют, что встраивание ИИ в экономику будет идти тем быстрее и безопаснее, чем раньше компании перестанут воспринимать безопасность и нормативку как «приложение к инновации» и начнут проектировать ИИ как часть управляемой, проверяемой и устойчивой системы. ■