Банк России

МАТЕРИАЛЫ ЗАСЕДАНИЯ ЭКСПЕРТНОГО СОВЕТА ПО РЕГУЛИРОВАНИЮ, МЕТОДОЛОГИИ ВНУТРЕННЕГО АУДИТА, ВНУТРЕННЕГО КОНТРОЛЯ И УПРАВЛЕНИЯ РИСКАМИ В БАНКЕ РОССИИ И ФИНАНСОВЫХ ОРГАНИЗАЦИЯХ ОТ 23 ИЮНЯ 2026 ГОДА

2026

ОГЛАВЛЕНИЕ

Обращение главного аудитора Банка России к читателям 3
Вступление 4
Аудит ИИ-моделей как драйвер зрелости СВА: взгляд со стороны регулятора (Т.Н. Мирук, Банк России) 10
Управление рисками искусственного интеллекта: правовые аспекты (В.В. Астанин, Университет Банка России) 22
Организация функции аудита моделей: 9 лет практики (О.В. Чистяков, ПАО Сбербанк) 27
Аудит жизненного цикла моделей: новые вызовы и развитие подходов (М.А. Мамалаева, АО «Альфа-Банк») 32
Управление рисками новых технологий: опыт Московской Биржи (С.В. Демидов, Группа «Московская Биржа») 35
Риски искусственного интеллекта: вызовы для внутреннего аудита и внутреннего контроля (А.В. Давыдова, ООО «Технологии доверия – консультирование») 39
Аудит искусственного интеллекта: вызов будущего (Г.Ю. Дельвиг, ПАО «Газпром нефть») 44
Искусственный интеллект в операционных процессах X5: основные риски (И.Н. Ворона, ПАО «Корпоративный центр ИКС 5») 49
Заключение 52
Экосистема экспертного совета 53
Глоссарий 55
Список сокращений 58

Материал подготовлен службой главного аудитора Банка России.

Ответственные за выпуск: Н.А. Станик, В.А. Щукин.

Комментарии, предложения и замечания можно направлять по адресу: expert.board@mail.cbr.ru.

107016, Москва, ул. Неглинная, 12, к. В

Официальный сайт Банка России: www.cbr.ru

ОБРАЩЕНИЕ ГЛАВНОГО АУДИТОРА БАНКА РОССИИ К ЧИТАТЕЛЯМ

Уважаемые коллеги!

Искусственный интеллект (ИИ) перестал быть исключительно технологической новацией. Сегодня он становится важной частью деятельности финансовых организаций, компаний реального сектора экономики и органов государственного управления, меняя подходы к принятию решений, организации бизнес-процессов и построению систем внутреннего контроля.

Одновременно расширяется круг вопросов, связанных с управлением модельными и ИИ-рисками. Разработка, внедрение и эксплуатация моделей требуют новых подходов к внутреннему аудиту, внутреннему контролю и управлению рисками, а также совершенствования корпоративного управления в целом.

Именно этим вопросам было посвящено очередное заседание Экспертного совета по регулированию, методологии внутреннего аудита, внутреннего контроля и управления рисками в Банке России и финансовых организациях.

Особенностью подготовки заседания стало внедрение нового формата работы Экспертного совета. Впервые каждому выступлению предшествовал тематический мастер-класс, в рамках которого авторы смогли представить свои подходы, обсудить практические вопросы с профессиональным сообществом и учесть результаты дискуссии при подготовке итоговых материалов. Подробнее об этом формате рассказано во вступлении к настоящему выпуску.

В этом дайджесте собраны материалы, отражающие как методологические подходы регулятора, так и практический опыт финансовых организаций, компаний реального сектора экономики, научного сообщества и ведущих российских экспертов. Несмотря на разнообразие рассматриваемых вопросов, все публикации объединяет общая цель – развитие современных подходов к управлению модельными и ИИ-рисками, основанных на эффективной системе внутреннего аудита, внутреннего контроля и управления рисками.

Надеюсь, что материалы выпуска будут полезны специалистам в области внутреннего аудита, внутреннего контроля и управления рисками, руководителям организаций, членам советов директоров и комитетов по аудиту, а также всем, кто занимается вопросами внедрения и безопасного использования ИИ-технологий.

В.П. Горегляд
Главный аудитор Банка России, председатель Экспертного совета по регулированию, методологии внутреннего аудита, внутреннего контроля и управления рисками в Банке России и финансовых организациях

ВСТУПЛЕНИЕ

Настоящий выпуск дайджеста подготовлен по итогам заседания Экспертного совета по регулированию, методологии внутреннего аудита, внутреннего контроля и управления рисками в Банке России и финансовых организациях, состоявшегося **23 июня 2026 года в Банке России**. Темой заседания стали вопросы управления модельными и ИИ-рисками, а также практика их внутреннего аудита и контроля.

Заседание прошло в очно-дистанционном формате и объединило около 500 участников. В его работе приняли участие специалисты Банка России, федеральных органов государственной власти, Национального банка Республики Беларусь, Национального банка Таджикистана, финансовых организаций, компаний реального сектора экономики, аудиторских и консалтинговых организаций, а также представители научного и экспертного сообщества.

Особенностью подготовки заседания стало внедрение нового формата работы Экспертного совета. Впервые выступлениям на заседании предшествовала серия тематических мастер-классов; в большинстве случаев их проводили сами авторы будущих докладов, а по отдельным темам – их коллеги, детально знающие материал. Каждый мастер-класс позволил заранее подробно рассмотреть заявленную тему, обсудить практические вопросы с участниками и получить профессиональную обратную связь еще до проведения заседания. Подробнее о новом формате работы рассказано в отдельном подразделе дайджеста.

Благодаря такому подходу материалы этого выпуска отражают не только авторские исследования и практический опыт, но и результаты профессионального обсуждения, состоявшегося в ходе мастер-классов и заседания Экспертного совета.

Дайджест объединяет публикации, посвященные различным аспектам управления модельными и ИИ-рисками: вопросам регулирования и надзорной практики, правовым аспектам применения ИИ, организации внутреннего аудита моделей, управлению их жизненным циклом, построению корпоративных фреймворков управления ИИ-рисками, а также практическому опыту внедрения ИИ в деятельность российских организаций.

Новый формат работы Экспертного совета

В 2026 году Экспертный совет приступил к реализации нового формата подготовки заседаний, ориентированного на более глубокую проработку рассматриваемых вопросов и расширение профессиональной дискуссии.

Подготовка каждого заседания включает несколько последовательных этапов:

- проведение тематических мастер-классов;
- обсуждение практических вопросов с профессиональным сообществом;
- выступления авторов на заседании Экспертного совета;
- публикация материалов в виде тематического дайджеста.

Каждый мастер-класс продолжительностью от полутора до двух часов проводится автором будущего доклада либо его коллегой, детально знающим соответствующий вопрос. Такой формат дает возможность рассмотреть тему значительно подробнее, чем позволяет регламент заседания, обсудить практические кейсы, ответить на вопросы участников и учесть их предложения при подготовке итоговой публикации.

Практика показала, что предварительное профессиональное обсуждение способствует более глубокой проработке материалов и делает итоговые публикации более содержательными. В результате настоящий дайджест отражает не только позицию авторов, но и выводы, сформировавшиеся в ходе экспертной дискуссии.

Благодарность авторам

Редакционная коллегия выражает искреннюю признательность авторам, подготовившим материалы и представившим доклады на заседании Экспертного совета.

В подготовке дайджеста приняли участие:

- Т.Н. Мирук – Банк России;
- В.В. Астанин – Университет Банка России;
- О.В. Чистяков – ПАО «Сбербанк»;
- Н.М. Мануйлова и М.А. Мамалаева – АО «Альфа-Банк»;
- С.В. Демидов – Группа «Московская Биржа»;
- А.В. Давыдова – ООО «Технологии Доверия – Консультирование»;
- Г.Ю. Дельвиг и Ю.П. Максимчук – ПАО «Газпром нефть»;
- И.Н. Ворона – ПАО «Корпоративный центр ИКС 5».

Публикации объединили методологические разработки, результаты исследований и практический опыт организаций финансового и реального секторов экономики. Благодаря этому читатели получают возможность познакомиться с различными подходами к управлению модельными и ИИ-рисками, к организации внутреннего аудита моделей и развитию современных систем внутреннего контроля.

Редакционная коллегия также благодарит всех участников заседания за активное участие в профессиональной дискуссии, обмен опытом и вклад в развитие современных подходов к внутреннему аудиту, внутреннему контролю и управлению рисками.

Благодарность ведущим мастер-классов

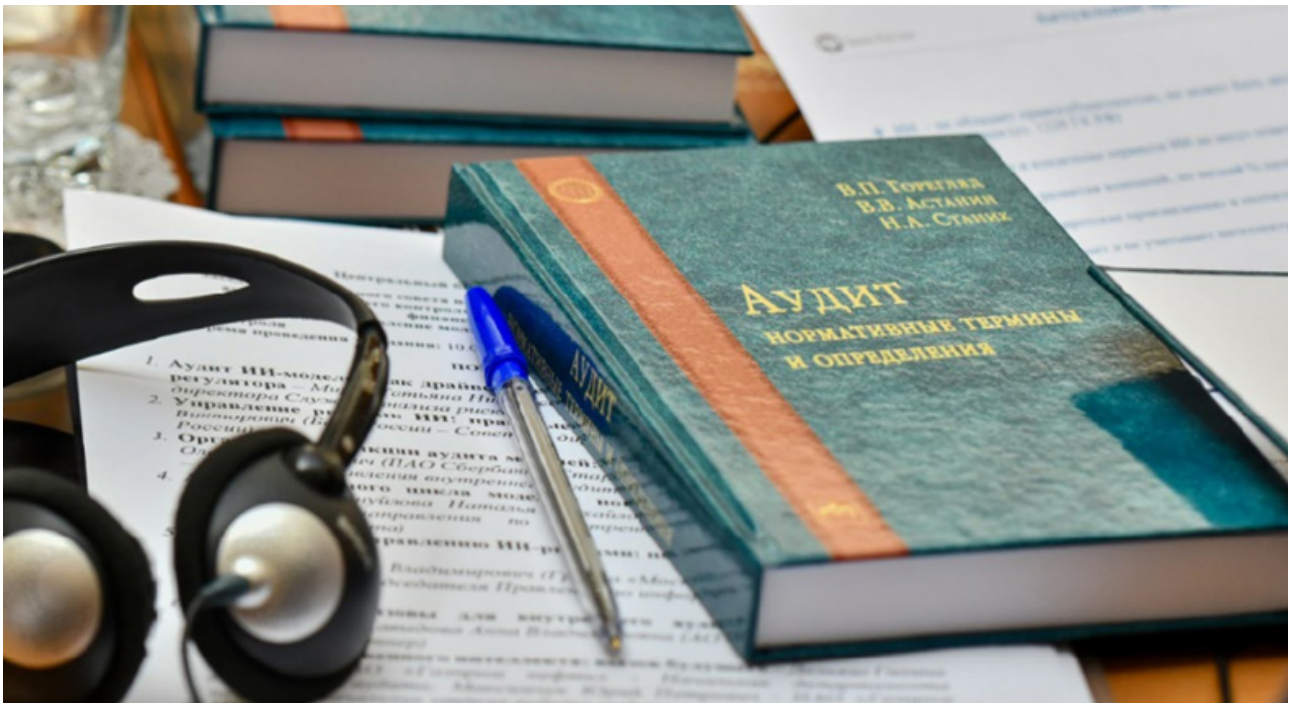
Отдельную благодарность редакционная коллегия выражает экспертам, которые провели тематические мастер-классы, предварившие доклады на заседании Экспертного совета. По ряду тем мастер-класса его проводили не сами докладчики, а их коллеги, вклад которых в подготовку содержательной профессиональной дискуссии не менее важен:

- Т.Н. Мирук, О.К. Васильева, Д.А. Жевага – Банк России;
- В.В. Астанин – Университет Банка России;
- В.Ю. Прокофьев – ПАО «Сбербанк»;
- М.А. Мамалаева – АО «Альфа-Банк»;
- С.В. Демидов – Группа «Московская Биржа»;
- А.В. Давыдова – ООО «Технологии Доверия – Консультирование»;
- Г.Ю. Дельвиг, Ю.П. Максимчук – ПАО «Газпром нефть»;
- И.Н. Ворона – ПАО «Корпоративный центр ИКС 5».

Редакционная коллегия благодарит всех ведущих мастер-классов за готовность заранее подробно представить свои подходы и открыто обсудить практические вопросы с профессиональным сообществом.

Презентация нового издания

Одним из событий заседания стала презентация книги «Аудит: нормативные термины и определения», подготовленной В.П. Гореглядом, В.В. Астаниным и Н.А. Станик и выпущенной издательством «ИНФРА-М» в 2026 году.



Представляя издание, В.П. Горегляд подчеркнул, что дальнейшее развитие профессии невозможно без формирования единого понятийного аппарата, распространения лучших практик и подготовки современных методических материалов, отражающих ее актуальные направления.

Выход издания стал одним из важных результатов работы Экспертного совета. Наряду с тематическими мастер-классами, заседаниями Экспертного совета и настоящим дайджестом, оно отражает стремление сообщества объединять лучшие практики и поддерживать единое понимание профессиональной терминологии.

В знак признательности за вклад в подготовку заседания и дайджеста экземпляры книги были вручены спикерам.

Итоги заседания Экспертного совета и краткий обзор докладов

В повестку заседания было включено девять докладов; восемь из них представлены в ходе заседания (доклад «ИИ: угрозы и возможности», заявленный Р.М. Гусейновым, не состоялся).

Каждому докладу предшествовал отдельный тематический мастер-класс – в большинстве случаев его проводил сам докладчик, а по отдельным темам – коллега докладчика (полный список – в подразделе «Благодарность ведущим мастер-классов»).

Ниже приведен краткий обзор представленных докладов.

АУДИТ ИИ-МОДЕЛЕЙ КАК ДРАЙВЕР ЗРЕЛОСТИ СВА: ВЗГЛЯД СО СТОРОНЫ РЕГУЛЯТОРА

В этих условиях повышается роль аудитора (особенно внутреннего!), который может оценить уровень риска, привносимый моделями ИИ, и дать руководству объективную картину.

Т.Н. Мирук, Банк России

Рассмотрена эволюция нормативной базы Банка России в части управления модельным риском, международная практика регулирования ИИ (ЕС, США, Китай, Индия, Россия), а также представлен опыт надзорного взаимодействия Банка России с организациями по вопросам валидации ML-моделей. Предложено сформировать рабочую группу при Экспертном совете для выработки методических рекомендаций по управлению модельными рисками и ИИ-рисками, аудиту моделей.

УПРАВЛЕНИЕ РИСКАМИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ПРАВОВЫЕ АСПЕКТЫ

ИИ не обладает правосубъектностью, не может быть автором и соавтором, а его контент не является произведением.

В.В. Астанин, Университет Банка России

Рассмотрены актуальные правовые позиции в отношении ИИ (отсутствие правосубъектности, вопросы охраноспособности сгенерированного контента), риски нарушения интеллектуальных прав при использовании ИИ работниками, судебная практика, а также положения законопроекта «Об основах государственного регулирования сфер применения технологий искусственного интеллекта в Российской Федерации». Представлена практическая карта рисков применения ИИ в организации и рекомендации по их минимизации.

ОРГАНИЗАЦИЯ ФУНКЦИИ АУДИТА МОДЕЛЕЙ: 9 ЛЕТ ПРАКТИКИ

Служба внутреннего аудита рассматривает модель как элемент бизнес-процесса и проводит оценку не только ее функционирования, но и потенциального влияния ограничений модели на клиентский опыт, финансовый результат и уровень рисков.

О.В. Чистяков, ПАО Сбербанк

Представлен опыт становления и развития функции аудита моделей в Службе внутреннего аудита ПАО Сбербанк с 2017 по 2026 год. Рассмотрены три составляющих организации Службы: люди (Data Scientists и эксперты в области аудита), технологии (вычислительные мощности, хранилища данных, собственное программное обеспечение) и фабрика знаний – механизм извлечения, систематизации и передачи экспертизы.

АУДИТ ЖИЗНЕННОГО ЦИКЛА МОДЕЛЕЙ: НОВЫЕ ВЫЗОВЫ И РАЗВИТИЕ ПОДХОДОВ (ПО МАТЕРИАЛАМ, ПОДГОТОВЛЕННЫМ СОВМЕСТНО С М.А. МАМАЛАЕВОЙ)

Нас интересует уже не только качество алгоритма. Нас интересует качество среды, в которой этот алгоритм живет, развивается и принимает решения.

Н.М. Мануйлова, АО «Альфа-Банк»

Показана эволюция аудита моделей: от проверки отчетов о разработке и валидации до системного аудита управляемости жизненного цикла модели «от идеи до вывода из эксплуатации». Выделены пять факторов, меняющих требования к аудиту моделей (количество моделей, скорость релизов, сложность, качество данных, распространение GenAI/LLM), и четыре направления развития аудита в ответ на эти вызовы.

ФРЕЙМВОРК ПО УПРАВЛЕНИЮ ИИ-РИСКАМИ: ПОДХОДЫ И ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ

ИИ становится самостоятельным источником риска, находящимся на стыке операционных, кибер-, этических и комплаенс-рисков.

С.В. Демидов, Группа «Московская Биржа»

Представлен интегрированный в общую систему управления рисками фреймворк управления ИИ-рисками Группы «Московская Биржа»: определение ИИ-риска как подвида нефинансового риска на стыке ИБ-, ИТ-, модельных, операционных, этических и регуляторных рисков; параметры присвоения уровня риска ИИ-системе по 12 критериям значимости и вероятности реализации риска; цикл управления ИИ-рисками на всех этапах жизненного цикла системы – от планирования до эксплуатации.

РИСКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ВЫЗОВЫ ДЛЯ ВНУТРЕННЕГО АУДИТА И ВНУТРЕННЕГО КОНТРОЛЯ

В условиях стремительной цифровизации ИИ перестает быть просто технологическим трендом и превращается в ключевой инструмент управления.

А.В. Давыдова, ООО «Технологии Доверия – Консультирование»

Проанализированы системные причины, по которым значительная часть ИИ-проектов не достигает заявленных целей: недостаточная бизнес-ценность, низкое качество данных, рост совокупной стоимости владения, неэффективное управление изменениями. Рассмотрен вопрос распределения ответственности за ошибки ИИ-инструментов между разработчиком, владельцем решения и руководством организации. Представлена адаптированная модель цифрового доверия для анализа ИИ-решений внутренними аудиторами.

АУДИТ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ВЫЗОВ БУДУЩЕГО

Нужен симбиоз человеческого скептицизма и машинного интеллекта.

Г.Ю. Дельвиг, ПАО «Газпром нефть»

Рассмотрена трансформация функции внутреннего аудита в условиях распространения ИИ-агентов: новые роли аудитора (бизнес-партнер, аналитик данных, архитектор доверия), необходимость создания изолированной технологической среды («песочницы») для тестирования новых моделей, а также библиотеки промптов и набора цифровых ассистентов для самой функции аудита. Сформулированы семь вопросов, которые аудитор должен задавать любой системе с ИИ, и отдельные вопросы для проверки автономных агентов.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ОПЕРАЦИОННЫХ ПРОЦЕССАХ X5: ОСНОВНЫЕ РИСКИ

На практике нельзя на 100% отследить, что и как делает ИИ-агент, но без контроля ИИ может превратиться в черный ящик.

И.Н. Ворона, ПАО «Корпоративный центр ИКС 5»

Представлен практический опыт внедрения ИИ в операционную деятельность розничной компании: прогнозирование спроса, персонализация предложений покупателям, проверка договоров аренды, компьютерное зрение для контроля выкладки товара и соблюдения гигиенических требований на производстве. Обозначены три основных риска использования ИИ (утечка конфиденциальных данных, риски вычислительной инфраструктуры, риск ошибочных действий ИИ-агентов из-за галлюцинаций) и меры по их митигации, включая размещение моделей в закрытом контуре и подход ИИ-аугментации.

По итогам обсуждения участники заседания отметили необходимость дальнейшего развития методологической базы аудита моделей и ИИ-систем, расширения профессиональных компетенций внутренних аудиторов в этой области и продолжения обмена практическим опытом в рамках Экспертного совета, в том числе через практику подготовительных тематических мастер-классов.

Банк России продолжит придерживаться риск-ориентированного и технологически нейтрального подхода с фокусом на мягкое регулирование управления модельным риском и ИИ-риском.

Т.Н. МИРУК
Заместитель руководителя Службы анализа рисков Банка России

АУДИТ ИИ-МОДЕЛЕЙ КАК ДРАЙВЕР ЗРЕЛОСТИ СВА: ВЗГЛЯД СО СТОРОНЫ РЕГУЛЯТОРА

Аннотация

Модельный риск как фактор неопределенности тесно связан с расширяющимся использованием статистических моделей и ИИ-моделей как в процессах принятия решений, так и в процессах поддержки текущей операционной деятельности. В последние годы увеличение вклада модельного риска в мировые финансовые кризисы, повсеместная цифровизация, внедрение генеративного ИИ, повышение сложности моделей подтверждают актуальность задачи контроля за модельным риском и ИИ-риском.

В связи с этим растет роль риск-менеджмента и аудита при оценке модельного риска и ИИ-риска – в первую очередь для контроля за полным жизненным циклом модели, независимым распределением функций при создании, валидации и мониторинге модели, оценке достаточности и документированности набора индикаторов и тестов, характеризующих качество модели.

Ключевые слова: модель, операционный риск, модельный риск, ИИ-риск, жизненный цикл модели, реестр моделей, ролевая матрица, аудит моделей.

Коды JEL: G18, G28, G40, O38, M42.

Понятие модельного риска и ИИ-риска. Нормативная база Банка России

В текущей нормативной базе Банка России модельный риск (включая ИИ-риск) рассматривается как вид операционного риска. Определения данных видов риска вводятся Указанием Банка России № 3624-У. Так, под **операционным риском** понимается риск возникновения прямых и непрямых потерь в результате несовершенства или ошибочных внутренних процессов кредитной организации, действий персонала и иных лиц, сбоев и недостатков информационных, технологических и иных систем, а также в результате реализации внешних событий. **Модельный риск** определяется внутри операционного риска как риск ошибок процессов разработки, проверки, адаптации, приемки, применения методик количественных и качественных моделей оценки активов, рисков и иных показателей, используемых в принятии управленческих решений. При этом отдельного определения ИИ-риска на текущий момент не сформулировано, он рассматривается как вид модельного риска без предъявления к нему специфических требований.

Поскольку модельный риск является частью операционного риска, все требования Положения Банка России № 716-П к управлению операционным риском, в частности требования по сбору данных о случаях реализации риска, распространяются и на модельный риск. Кроме того, Положение Банка России № 716-П выделяет специфические требования к управлению модельным риском, но они обязательны для исполнения только крупными кредитными организациями (с размером активов 500 млрд рублей и более).

Стоит упомянуть о наличии российской надзорной практики по методологии разработки, валидации и мониторинге моделей, используемых для расчета величины кредитного риска с использованием подхода на основе внутренних рейтингов (Положение Банка России № 845-П), и моделей, оценивающих доход заемщика для расчета показателя долговой нагрузки (Указание Банка России № 6579-У).

В 2025 году Банк России выпустил «Кодекс этики в сфере разработки и применения ИИ на финансовом рынке», который закладывает основу мягкого регулирования ИИ на финансовом рынке, рекомендует финансовым организациям управлять ИИ-рисками в рамках общей системы управления рисками организации.

16.06.2026 Банк России выпустил Методические рекомендации № 3-МР по обеспечению информационной безопасности при разработке и применении искусственного интеллекта на финансовом рынке, которые фактически формируют новый стандарт корпоративного управления ИИ-рисками, переводя эту задачу из технической области в плоскость стратегического управления.

В ближайшее время в развитие темы управления модельным риском и ИИ-риском Банк России продолжит придерживаться риск-ориентированного и технологически нейтрального подхода с фокусом на мягкое регулирование, о чем говорится в докладе для общественных консультаций Банка России «Применение искусственного интеллекта на финансовом рынке». Планируется анализировать собранную поднадзорными организациями статистику в части потерь от модельного риска и ИИ-риска, осуществлять комплексный ежегодный мониторинг соблюдения Кодекса этики, формировать обзоры лучших практик его соблюдения. В результате будет возможно рассмотреть вопрос о необходимости разработки отдельного руководства по управлению ИИ-рисками.

Мировая практика

В разных юрисдикциях механизмы управления модельным риском и ИИ-риском базируются на различных подходах. Однако можно сделать общий вывод, что государства в целом озабочены вопросами создания регуляторной среды, при этом наблюдается тенденция к более мягкому регулированию у стран-лидеров в гонке за ИИ.

ЕС

Осуществляется жесткое риск-ориентированное регулирование, включающее запрет на неприемлемые риски, регистрацию высокорисковых систем в специальной базе, требования к качеству данных и человеческому надзору. Выявленные нарушения могут привести к крупным штрафам.

Казахстан

Создаются гибкие и привлекательные для технологического бизнеса условия внедрения ИИ: разработка первого в ЕАЭС закона «Об искусственном интеллекте» и формирование Министерства искусственного интеллекта с акцентом не на запретах, а на создании конкурентной среды и предоставлении бизнесу большего пространства для маневра по внедрению передовых технологий.

Китай

Осуществляются государственный контроль и поддержка инноваций. Вводятся обязательная регистрация и предзапуск, оценка алгоритмов, требования к контенту в части непротиворечия традиционным ценностям, строгие правила для генеративного ИИ. При этом поддерживаются национальные чемпионы в области ИИ.

США

Принят децентрализованный, инновационно ориентированный подход: регулирование определяется на уровне отдельных штатов, сделаны акценты на саморегулировании и отраслевых стандартах.

Индия

Нет отдельного закона о регулировании ИИ, осуществляется «легкое» регулирование, делается ставка на баланс между инновациями и безопасностью. Ключевой документ – Руководящие принципы управления ИИ, включающие «семь сутр» (доверие, человекоцентричность, инновации, справедливость, подотчетность, понятность; безопасность и устойчивость).

Значимость ИИ-риска

Внедрение ИИ усиливает и видоизменяет традиционные риски, а также создает принципиально новые:

- зависимость ИИ от сложной цепочки технологий (облачные сервисы, внешние поставщики данных) расширяет возможность внешних атак, а также «закладок» в коды программ, поддерживающих ИИ;
- интеграция ИИ с операционными системами может привести к техническим сбоям и масштабным нарушениям работы. Также возрастают требования к вычислительным мощностям и системам электропитания, которые поддерживают функционирование ИИ;
- вследствие того что ИИ обучен на огромных массивах неизвестных данных, его алгоритмы могут перенимать и усиливать предвзятость систем, что может привести к дискриминации, судебным искам и штрафам;
- впервые, в отличие от традиционных моделей и алгоритмов, человек не может полностью проанализировать логику работы ИИ и однозначно предсказать результат его работы;
- генеративные модели могут выдавать галлюцинации как правдоподобную, но совершенно неверную информацию, что может привести к неверным решениям;

- ИИ-агенты могут не просто советовать, а самостоятельно действовать в системах, и ошибка в действиях такого агента – это прямой операционный сбой;
- проблема авторского права – документы, на основании которых учится ИИ-модель, заведомо имеют правообладателя. При этом модель, обучаясь на этих документах, дает ответы, базирующиеся на них.

Остается открытым вопрос – кто именно несет юридическую и моральную ответственность за все возможные негативные последствия работы модели?

В докладе «Отчет об искусственном интеллекте и системных рисках», выпущенном в декабре 2025 года Европейской комиссией по системным рискам, сделан вывод: быстрое внедрение ИИ в финансовую систему может значительно усилить системные риски, особенно если оставить ИИ без должного контроля. Как следствие, государствам и регуляторам необходимо действовать на опережение.

Основные вызовы:
- монополизация и высокий порог входа – рынок контролируется малым количеством игроков. В случае сбоя крупного провайдера (например, поставщика модели или облачной платформы) это мгновенно парализует деятельность тысяч финансовых организаций;
- однородность моделей – все поставщики используют схожие алгоритмы. Новый вызов, который не учтен в моделях, может привести к некорректному срабатыванию алгоритмов;
- проблема с контролем (блэк-бокс) – сложность систем затрудняет выявление ошибок, модели не отвечают понятию интерпретируемости, а показатели качества моделей не всегда ясны как самим компаниям, так и надзорным органам. При этом появился новый тип ошибок – галлюцинации моделей, которые сложно выявить без глубокого понимания предметной области;
- чрезмерное доверие – все чаще решения и выводы доверяются ИИ, люди полагаются на ИИ в кризисных ситуациях, повышая уязвимость системы к ошибкам.

В этих условиях растет роль аудитора (особенно внутреннего), который может оценить существенность угроз, привносимых ИИ-моделями, и дать руководству объективную картину управления рисками.

Не стоит забывать и о основополагающих принципах модельного риска, сформулированных на регуляторном уровне применительно к финансовому рынку еще в начале 2000-х годов:
- все финансовые модели подвержены ошибкам;
- ключевые процедуры в рамках указанной деятельности должны сопровождаться мероприятиями по идентификации и устранению ошибок;
- эти мероприятия должны выполняться квалифицированным и независимым персоналом.

Ключевой момент – идентификация модели

Основной проблемой при определении модельного риска является множественность и несогласованность терминов, описывающих область моделирования. Это и типы моделей, и соответствующие методы, и достаточно размытые границы в терминологии (Data Science (DS) / Machine Learning (ML) / Artifical Intelligence (AI) / Deep Learning (DL)) в рамках применения моделей.

Под моделью могут пониматься математическая формула или алгоритм, основанный на экспертных предположениях, нейронная сеть, обученная на сложных наборах данных (датасетах), или генеративная модель, обученная на терабайтах видео- и текстовой информации.

Особенно это важно при разработке внутренних методологических документов финансовой организации, так как именно на этом основывается дальнейшая детализация процесса управления модельным риском и ИИ-риском.

Наиболее общепринятым в универсальном значении в мировой практике является определение модели в стандартах ISO (10303-1). В руководящем письме SR 11-7 ФРС США дает определение модели в финансах: *«количественный метод, система или подход, в котором применяются статистические, экономические, финансовые или математические теории, приемы и предположения для переработки входных данных в количественные или качественные оценки»*. Дальнейшая детализация этого понятия расширяется на уровне надзорного законодательства или в самих финансовых организациях.

Модели, отвечающие такому определению, могут использоваться для анализа бизнес-стратегий, информирования о бизнес-решениях, выявления и измерения рисков, оценки рисков, финансовых инструментов или позиций, проведения стресс-тестирования, оценки достаточности капитала, управления активами клиента, комплаенса соблюдения внутренних лимитов на операции, поддержания контрольных функций управляющего аппарата банка, соблюдения требований финансовой или нормативной отчетности и публичного раскрытия информации. Определение модели также охватывает количественные подходы, исходные данные которых являются частично или полностью качественными или основанными на экспертных оценках при условии, что выходные данные модели носят количественный характер.

В российском законодательстве однозначного универсального определения модели нет. Его содержание зависит от сферы применения: регулирование рисков (Банк России), инвестиционное проектирование (Минэкономразвития) или финансовый учет. Федеральный закон № 86-ФЗ и Положение Банка России № 845-П устанавливают требования к моделям и задачи их применения. В Положении Банка России № 716-П понятие модели неявно включено в понятие модельного риска: модельный риск как ошибка процессов разработки, проверки, адаптации, приемки, применения методик количественных и качественных моделей оценки активов, рисков и иных показателей, используемых в принятии управленческих решений.

В проекте федерального закона Российской Федерации «Об основах государственного регулирования сфер применения технологий искусственного интеллекта» говорится: *«Искусственный интеллект – комплекс технологических решений, позволяющих имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма), и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их»*.

На практике любая модель – это упрощение действительности, следовательно, она основывается на предположениях. Предположения могут относиться к характеристикам исходных данных, взаимосвязям между макроэкономическими факторами, прогнозам и возможным выводам. Главное требование к предположениям в рамках одной модели – стабильность. Если предположения будут меняться постоянно, то использование инструмента, который на них опирается, будет каждый раз новым и второе условие о повторном использовании не будет выполняться.

Реестр моделей

Для решения задач как оценки рисков моделей, так и последующего аудита моделей основным инструментом становится реестр моделей, включающий в себя полную информацию об имеющихся в организации моделях, в том числе их влиянии на финансовый результат и организационную структуру.

Например, структура реестра моделей может содержать информацию об идентификационных данных модели, качественных и количественных параметрах, определяющих основные показатели модели, способах и условиях применения модели, лицах, ответственных за разработку, валидацию и мониторинг модели, требования к информационной безопасности, программный код модели, описание используемых источников данных и так далее.

Подобная инвентаризация позволит:
- оценить интенсивность применения моделей различными подразделениями;
- проанализировать взаимосвязь моделей – например, когда данные одной модели являются входящими данными для других моделей;
- выявить всех участников модельного процесса или отсутствие ключевых участников в жизненном цикле конкретной модели;
- исключить дублирование моделей или функций для конкретных моделей.

При этом ведение реестра моделей обладает и явным экономическим эффектом – он позволяет минимизировать дублирование при разработке однотипных моделей в разных подразделениях и, соответственно, оптимизировать операционные затраты.

В рамках аудита моделей целесообразно придерживаться риск-ориентированного подхода. Организация должна определить критерии оценки уровня существенности модельного риска, включая критерии оценки потерь, количественной и качественной оценки вероятности реализации модельного риска в условиях текущего процесса, например:
- высокий риск – воздействие на критическую инфраструктуру организации, существенное влияние на финансовую устойчивость или вероятность нарушения этических норм;
- ограниченный риск – влияние на финансовый эффект отдельных продуктов или сервисов;
- минимальный риск – возможны незначительные финансовые потери при некорректно принятом решении.

Сведения об уровне существенности риска модели необходимо также отразить в реестре моделей.

Три кита аудита модельного риска

При аудите модельного риска и ИИ-риска необходима оценка наличия и качества реализации трех китов модельного риска:
1. Реализован полный жизненный цикл модели (ModelOps) – процесс построения модели от момента возникновения бизнес-идеи до внедрения, последующего мониторинга и вывода из эксплуатации модели (списания) при необходимости.
2. Существует независимое распределение функций при создании, валидации и мониторинге модели – распределение зон ответственности и ролей повышает управляемость процесса в целом, позволяет более эффективно оценивать его оптимальность и результативность, формировать обоснованную внутреннюю отчетность для топ-менеджмента организации.
3. Проводится оценка достаточности и документированности набора индикаторов и тестов, характеризующих качество модели, – развитие корпоративной культуры жизненного цикла моделей, которая предполагает документирование и протоколирование всех действий по созданию, валидации, применению и мониторингу моделей.

Остановимся на каждом из этих элементов подробнее.

1. MODEL RISK-MANAGEMENT И ЖИЗНЕННЫЙ ЦИКЛ МОДЕЛИ

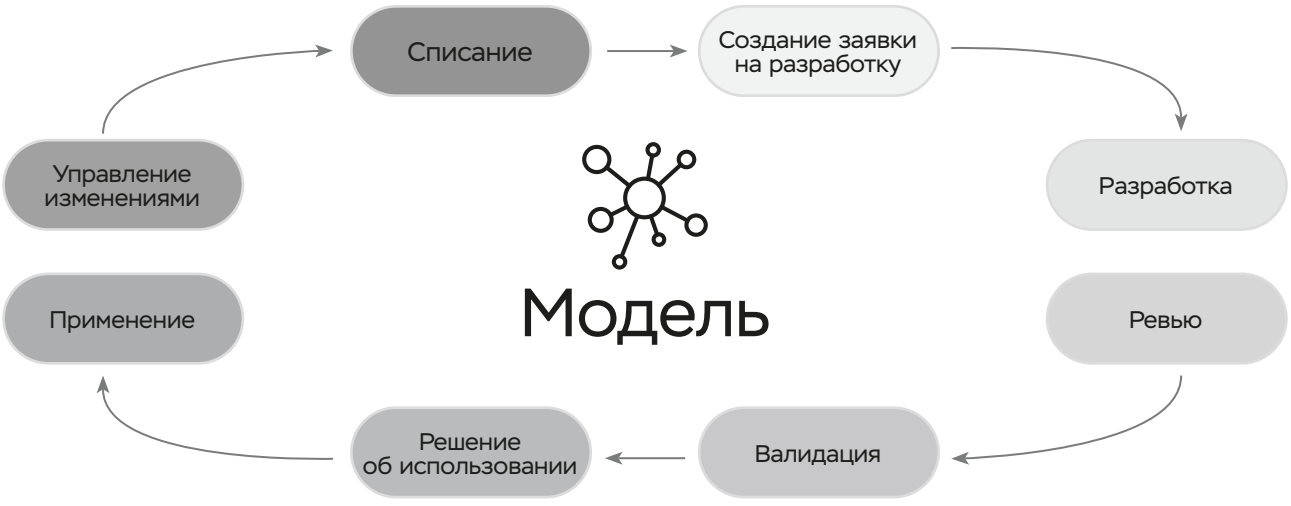
Ключевым аспектом управления модельным риском является синхронизация двух процессов – жизненного цикла моделей и управления модельным риском. Жизненный цикл моделей – это процесс построения модели от момента возникновения бизнес-идеи до внедрения, последующего мониторинга и вывода из эксплуатации модели (списания) при необходимости.

Во многих организациях можно столкнуться с тем, что различные подразделения одномоментно разрабатывают аналитические модели, не имея каких-либо формализованных или стандартизированных процессов хранения, внедрения и проверки. При этом также может отсутствовать (частично или полностью) документация о том, кому принадлежит модель, кем и для каких целей она применяется.

Без данной информации защитить использование модели перед регулятором / внешним проверяющим органом / внутренним аудитом становится практически невозможно. Также в отсутствие систематизированного процесса управления моделями повышаются затраты ресурсов на разработку, внедрение и поддержание их внутри организации. Лучшим решением в таких условиях будет детализация и формализация всех этапов жизненного цикла моделей.

В последнее время для описания жизненного цикла моделей используют ModelOps – набор практик и технологий для упрощения внедрения аналитических моделей в промышленную эксплуатацию. Практики ModelOps направлены прежде всего на упрощение взаимодействия аналитиков данных и ИТ-служб посредством максимальной автоматизации технических процессов внедрения моделей, а также жизненного цикла моделей и процесса управления модельным риском.

В качестве примера можно привести следующее описание жизненного цикла модели:



Создание заявки на разработку

Владелец модели формирует заявку на создание модели. В ней он классифицирует будущую модель (например, в разрезе типа продукта и типа модели), описывает, для чего и зачем она нужна.

Разработка модели

На этом этапе, помимо подготовки данных, происходит выбор наиболее подходящих технологии и метода для модели. Далее производится реализация кода модели и сопутствующих инструментов (например, кода для подгрузки данных), а также документирование всех этапов разработки, включая обоснование выбранного метода. По завершении этапа модель проходит первичные этапы тестирования и проверки.

Ревью (критический анализ / рецензирование) модели

Лидер команды разработчиков анализирует и подтверждает информацию о реализованной модели, в том числе последовательность и достаточность документации. При отсутствии требуемых данных лидер команды разработки обязан уведомить и запросить детализацию у разработчиков и только после этого передать модель группе валидаторов на оценку.

Валидация модели

Валидаторы работают уже с финальной версией модели и данных и не имеют доступа к рабочему пространству разработчиков. На этом этапе анализируются методологическая корректность, полнота документации и соответствие изначально поставленной задаче. Также валидаторы проводят количественные тесты для оценки производительности и стабильности модели. На основании результатов готовится отчет о проделанной работе.

Принятие решения о вводе модели в эксплуатацию

Коллегиальный орган или группа по управлению модельным риском, основываясь на результатах этапа валидации и дополнительном собственном анализе, принимают финальное решение – можно ли выводить модель в промышленную эксплуатацию или ее необходимо направить на доработку.

Промышленная эксплуатация модели

На этом этапе бизнес-пользователи применяют модель в рамках своих процессов и принимают стратегические решения на основании полученных результатов. В это же время группа валидаторов проводит периодический мониторинг модели и фиксирует все выявленные недостатки. На каждый недостаток создаются заявки на изменение модели, которые обрабатывают разработчики.

Списание модели

Помимо валидаторов, контролировать модель продолжает группа по управлению модельным риском, которая следит за количеством недостатков и статусом их закрытия. По общеустановленному правилу она отслеживает критическую границу количества незакрытых недостатков и степень их влияния. В случае преодоления этой границы группа инициирует процесс списания модели в архив по причине высокого модельного риска.

Кроме рассмотренного жизненного цикла модели, возможны и другие вариации. При этом количество этапов может зависеть от типа модели или другого фактора.

2. РОЛЕВАЯ МОДЕЛЬ И РАСПРЕДЕЛЕНИЕ ФУНКЦИЙ

Основной проблемой, которая приводит к реализации модельного риска, может стать несогласованность действий участников процесса и нарушение сроков получения конечного результата. Кроме прямого влияния на финансовый результат деятельности бизнес-заказчика, эта несогласованность может привести к повышенному модельному риску и возникновению претензий со стороны надзорного органа.

В качестве примера можно привести следующее распределение ролей:

Роль	Описание функционала
Инициатор/заказчик	Владелец основного бизнес-процесса, в который встраивается модель
Методолог/куратор	Менеджер проекта и технолог, который ведет проект по внедрению модели и разбирается как в принципах работы модели, так и в бизнес-процессе, в который встраивается модель. Является центром компетенций, обладающим полным пониманием методологии процесса, использующего модель
Разработчик модели	Может иметь недостаточно компетенций для понимания бизнес-процесса, но должен понимать математические принципы разрабатываемой модели
Data Engineer	Инженер по данным, который отвечает за сбор данных и создание витрины для разработки модели
Валидатор	Независимый от Data Scientist сотрудник, проводящий валидацию модели
ИТ-архитектор	Сотрудник ИТ-подразделения, который отвечает за архитектурное решение по имплементации модели в ИТ-системы
ИТ-разработчик	Сотрудник, осуществляющий внедрение модели в операционный процесс
Аудитор	Оценивает и подтверждает, что процесс функционирует корректно, роли распределены, документирование обеспечено, набор индикаторов достаточен

Детализация в части распределения зон ответственности и определения ролей повысит управляемость процессом в целом, позволит более эффективно оценивать его оптимальность и результативность, степень достижения целей его участниками на каждом этапе жизненного цикла модели, формировать обоснованную внутреннюю отчетность для топ-менеджмента организации.

3. ДОСТАТОЧНОСТЬ И ДОКУМЕНТИРОВАНИЕ НАБОРА ИНДИКАТОРОВ И ТЕСТОВ

Качественный аудит ИИ-модели – это многомерный подход, который сочетает в себе:

- количественный анализ с использованием набора технических, операционных и бизнес-метрик;
- качественный аудит для проверки полноты и обоснованности документации, подтверждающей, что модель создана и эксплуатируется прозрачно, ответственно и согласно собственной методологии и регуляторным нормам.

Документирование может содержать следующую информацию:

- аудит-трейл результатов отработки моделей и результирующих принятых решений – обычно это системные и пользовательские протоколы, которые могут описать процесс функционирования модели, а также помочь в случаях, когда модель слабоинтерпретируема;
- метрики качества ML-моделей – показатели, характеризующие точность и корректность работы модели (например, Gini / ROC-AUC – способность модели различать классы, MAE/RMSE – ошибка предсказания и так далее);
- метрики качества ИИ-моделей: Hallucination Rate – частота генерации фактически неверной информации, Answer Relevancy – соответствие ответа заданному вопросу, Correctness/ Factuality – соответствие ответа проверенным источникам, Coherence & Fluence – логичность и естественность текста;

- этические и регуляторные индикаторы: Bias/Fairness – дискриминация по демографическим признакам, Explainability – понятность решений для человека, Toxicity/Safety – наличие вредоносного/опасного контента, Privacy Compliance – соответствие требованиям защиты данных, Auditability – возможность реконструкции решений;
- непосредственно сами документы: политика по управлению модельным риском и ИИ-риском, реестр моделей (электронный), регламент взаимодействия, регламент мониторинга, матрица ролей, отчет о разработке модели, отчет о валидации модели, отчет о мониторинге и так далее.

Процесс аудита моделей

Аудит ИИ-моделей – это комплексный процесс, который выходит за рамки простой проверки точности работы модели. Он охватывает юридические, этические, операционные и технические аспекты.

Ниже приведена примерная пошаговая методология проведения полноценного аудита ИИ-моделей:

Шаг 1. Инициация и планирование аудита

На этом этапе закладывается основа всего последующего аудита: аудитор определяет контекст, границы и цели проверки, прежде чем переходить к анализу самой модели:

- перед тем как смотреть на код или данные, аудиторы должны понять контекст использования ИИ;
- классификация риска – определение уровня риска системы – высокий, ограниченный или минимальный (от этого зависит глубина аудита);
- определение бизнес-цели – какую проблему решает модель и каковы метрики успеха бизнеса?
- выявление заинтересованных сторон и участников процесса – кто влияет на модель, кто принимает решения на ее основе и на кого эти решения воздействуют (конечные пользователи, лица, принимающие решения, разработчики)?
- формирование команды – в команду аудита должны входить не только профессиональные аудиторы, но и Data Scientists, юристы, специалисты по ИИ, бизнес-аналитики, эксперты по информационной безопасности и так далее.

Шаг 2. Аудит системы управления

Проверка того, как процессы разработки и внедрения ИИ осуществляются внутри организации:

- инвентаризация моделей – есть ли реестр моделей, где зафиксированы ограничения, предполагаемое использование, метрики и дата обучения?
- роли и ответственность – закреплены ли ответственные за разработку, валидацию, мониторинг моделей, какое законодательство регламентирует их деятельность и соблюдаются ли его требования, как осуществляется контроль доступа к исходным данным, результатам работы модели, документации?

Шаг 3. Аудит данных

Данные – это фундамент ИИ. Для того чтобы исключить ошибки, которые приводят к катастрофическим последствиям («мусор на входе – мусор на выходе»), необходимы:

- качество и репрезентативность – проверка выборок на пропуски, дубликаты и аномалии, а также проверка того, отражает ли обучающая выборка реальную популяцию, к которой будет применяться модель;

- выявление смещений – анализ исторических данных на наличие аномалий (недостоверные данные) или смещений (некорректное распределение) и предубеждений (например, по полу, возрасту, региону);
- конфиденциальность – проверка методов анонимизации и псевдонимизации: исключено ли использование чувствительных персональных данных без явного согласия и кто имеет доступ к решениям модели, которые содержат чувствительные данные?
- наличие дрейфа – возможность отследить происхождение каждого датасета, его трансформации и версии.

Шаг 4. Технический аудит модели

Самый глубокий этап, на котором оценивается математический и алгоритмический аппарат:
- обоснованность выбранного матаппарата – в качестве матаппарата для модели может выступать как простейшая линейная регрессия, так и генеративная ИИ-модель. Насколько оправдан (в частности, экономически) выбранный метод?
- оценка метрик качества – проверка метрик (Accuracy, Precision, Recall, F1, ROC-AUC) не только в среднем по популяции, но и на отдельных сегментах данных;
- объяснимость и интерпретируемость – использование методов SHAP, LIME или Counterfactuals для проверки того, почему модель принимает те или иные решения. Модель не должна быть абсолютно черным ящиком для критически важных решений (например, для отказа в кредите);
- устойчивость – стресс-тестирование модели: как она реагирует на шум в данных и проводится ли тестирование на уязвимость к состязательным атакам и отравлению данных?

Шаг 5. Аудит внедрения и эксплуатации

Модель может быть идеальной при разработке, но сломаться в эксплуатации. Избежать этого помогут:
- мониторинг дрейфа – настроены ли триггеры на изменение распределения входных данных и изменение взаимосвязи между данными и целевой переменной?
- влияние человеческого фактора на решение – предусмотрена ли возможность вмешательства оператора, есть ли механизмы ручного переопределения ИИ-решений в спорных случаях?
- план отката – что произойдет, если модель начнет генерировать критические ошибки, и есть ли автоматический откат на предыдущую (эвристическую или более старую) версию?
- настройка автоматических триггеров – существуют ли механизмы, которые автоматически блокируют работу модели или отправляют ее на принудительный пересмотр, если показатели дрейфа данных или бизнес-метрик выходят за заданные рамки?
- наличие альтернативной ветви бизнес-процесса – возможно ли переключение на ручной режим работы в случае отказа модели и с какими рисками это переключение сопряжено?
- аудит моделей как непрерывный процесс, а не разовое событие – как часто должен аудироваться весь процесс разработки моделей, а также конкретные модели и какие существуют механизмы внепланового аудита?

Пример практической реализации надзорной деятельности в Банке России

В качестве примера практического опыта надзорной деятельности Банка России в части регулирования модельного риска можно привести кейс по надзорному взаимодействию с организациями, использующими в своей деятельности ML-модели. Также у Банка России есть опыт внутренней разработки и валидации собственных моделей, ведения реестра моделей.

За основу были взяты практики ModelOps (или ModelGoverment), а также использован опыт валидации ПВР-моделей крупнейших банков. Подразделения Банка России, обладающие в том числе опытом оценки внешних моделей, уделяют большое внимание разработке собственных моделей. При этом одним из ключевых фокусов внимания является оценка и минимизация модельных рисков. Такой подход позволяет, с одной стороны, повышать качество работы подразделений, а с другой – получать практический опыт и формировать наиболее оптимальные подходы к управлению модельным риском. И этот опыт, по нашему мнению, может быть использован как для расширения внедрения стандартов работы с моделями внутри Банка России, так и для интеграции в регулирование для всего рынка.

Предложения по развитию нормативной практики

С целью объединения накопленного опыта Банка России и финансовых организаций, а также распространения лучших практик предлагаем организовать рабочую группу при Экспертном совете по регулированию, методологии внутреннего аудита, внутреннего контроля и управления рисками в Банке России и финансовых организациях. Задачей рабочей группы будет формирование подходов и методических рекомендаций к управлению модельными рисками и ИИ-рисками, аудиту моделей.

Список источников

1. Федеральный закон от 10.07.2002 № 86-ФЗ (ред. от 10.06.2026) «О Центральном банке Российской Федерации (Банке России)».
2. Информационное письмо Банка России от 09.07.2025 № ИН-016-13/91 «О кодексе этики в сфере разработки и применения искусственного интеллекта на финансовом рынке».
3. Методические рекомендации Банка России от 16.06.2026 № 3-МР по обеспечению информационной безопасности при разработке и применении искусственного интеллекта на финансовом рынке.
4. Положение Банка России от 08.04.2020 № 716-П «О требованиях к системе управления операционным риском в кредитной организации и банковской группе».
5. Положение Банка России от 02.11.2024 № 845-П «О порядке расчета величины кредитного риска банками с применением банковских методик управления кредитным риском и моделей количественной оценки кредитного риска».
6. Указание Банка России от 15.04.2015 № 3624-У «О требованиях к системе управления рисками и капиталом кредитной организации и банковской группы».
7. Указание Банка России от 16.10.2023 № 6579-У «О требованиях к порядку расчета суммы величин среднемесячных платежей и среднемесячного дохода заемщика».
8. ГОСТ Р 59277-2020 «Системы искусственного интеллекта. Классификация систем искусственного интеллекта».
9. Глоссарий терминов Комитета по системам платежей и расчетам (The Committee on Payment and Settlement Systems) (CPSS) Банка международных расчетов.
10. Доклад Технического комитета 164 и Высшей школы экономики «О национальной системе оценки соответствия технологий искусственного интеллекта требованиям в области функциональной корректности и безопасности» на форуме «Технологии и безопасность» 13.02.2024.
11. Руководящее письмо SR 11-7 Совета Управляющих Федеральной резервной системы и Управления валютного контролера «Руководство для регуляторов по управлению модельным риском» (SR Letter 11-7 «Supervisory Guidance on Model Risk Management» of the Board of Governors of the Federal Reserve System (Federal Reserve Board) и Office of the Comptroller of the Currency (OCC)) от 04.04.2011. (OCC-Bulletin 2011-12).

Ознакомиться с презентацией

УПРАВЛЕНИЕ РИСКАМИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ПРАВОВЫЕ АСПЕКТЫ

Составитель запроса к ИИ не является субъектом договора авторского заказа (статья 1288 ГК РФ), равно как и ИИ (его разработчик или владелец сервиса), по причине отсутствия субъективированного исполнителя создания произведения, в какой бы форме оно не было.

В.В. АСТАНИН
Доктор юридических наук, профессор, советник директора Университета Банка России

Аннотация

В статье освещаются юридически значимые аспекты применения ИИ в операционной деятельности организаций в контексте рисков утраты нематериальных активов, нарушения интеллектуальных и личных неимущественных прав, гражданско-правовых споров по вопросам распределения исключительных прав в отношении созданных объектов интеллектуальной собственности при помощи генеративных моделей искусственного разума. Рассматриваются феномен сиротских произведений, генерируемых ИИ в рисках недоступного правообладателя, потенциал охраноспособности созданных при помощи ИИ служебных материалов, вопросы патентования ИИ-инструментов. Приводятся судебные решения по оценке синтетического создания интеллектуальных объектов, их правомерного оборота в бизнес-процессах компаний. На примерах зарубежной и отечественной практики иллюстрируются прецеденты, обусловливающие изменение (прекращение) трудовых правоотношений с работниками и обязанности работодателя в связи с внедрением ИИ в реализацию трудовых функций. Анализируются положения законопроекта «Об основах государственного регулирования сфер применения технологий искусственного интеллекта в Российской Федерации» в части прогноза применения его норм, способных кардинально влиять на практику деловой активности компаний. Предлагаются конкретные меры нормативного, организационного, прикладного значения для внедрения в организациях с целью исключения правовых рисков, связанных с ИИ.

Ключевые слова: искусственный интеллект, интеллектуальная собственность, нейросети, правообладатель, исключительные права, интеллектуальные права, сиротские произведения, охраноспособность, гражданско-правовые споры, правовые риски, юридическая ответственность, трудовые правоотношения, законодательные предложения, кадровое администрирование, зарубежный опыт.

Коды JEL: K15, K31, K42, M54, O34.

Позиции ИИ в контексте отношений, связанных с оборотом синтетических произведений, созданных без творческих усилий человека, их охраноспособности, защиты интересов правообладателей контента, используемого для обучения искусственного разума, получают справедливые оценки неполноты и сложностей правового регулирования [2, 4].

С позиции фундаментальных основ авторского права неопровержимы критерии, в соответствии с которым ИИ не может быть создателем творческих произведений, а также признаваться соавтором.

В соответствии с нормами российского законодательства (статьи 1257 и 1258 ГК РФ) автором и, соответственно, соавтором произведения науки, литературы или искусства признается гражданин, творческим трудом которого оно создано. При этом правовые эксперименты по наделению правосубъектностью ИИ-агентов в зарубежной юрисдикции оказались не состоятельны в признании невозможности определить полноценность синтетического носителя разума (даже несмотря на присвоение ему почетного гражданства в Саудовской Аравии с именем София) [6] в категориях сознания, воли, возможностей распоряжаться правами, нести самостоятельную юридическую ответственность.

В контексте основополагающих канонов права интеллектуальной собственности, несмотря на генерируемый контент, ИИ также не является автором результата интеллектуальной деятельности, в отличие от гражданина, творческим трудом которого создается таковой (статья 1228 ГК РФ). Иными словами, в результате продукции ИИ недостаточно творческого вклада (труда) человека. Резонно может звучать замечание о том, что составитель запроса на создание текстового, графического или иного объекта, получение какой-либо информации в объективной форме проявляет творческий труд, вносит личный вклад в создание ожидаемого результата путем составления промпта. Между тем составитель запроса (промпта) не является субъектом договора авторского заказа (статья 1288 ГК РФ), равно как и ИИ (его разработчик или владелец сервиса), по причине отсутствия субъективированного исполнителя создания произведения, в какой бы форме оно не было.

Приведенные позиции демонстрируют не только картину отсутствия правового статуса ИИ, но и актуализируют проблемы появления с его помощью сиротских произведений, автор которых не известен. Исходя из концепции правового регулирования юридической судьбы и гражданско-правового оборота сиротских произведений, предполагается, что таковые охраняются авторским правом, но самого автора идентифицировать невозможно. В связи с этим использование сиротского произведения влечет за собой определенный риск, поскольку автор или правообладатель может объявиться и обратиться в суд по факту нарушения авторских и смежных прав. В отношении произведений, созданных при помощи ИИ, операциональное определение «сиротское произведение» можно применять по такой аналогии, поскольку материал обучения ИИ может содержать контент, у которого есть правообладатель.

Действующие с 2024 года законодательные нормы регулирования отношений (содержащие поправки к ГК РФ), связанные с орфанными произведениями (от англ. «orphan» – сирота) [1], не могут быть распространены на созданные при помощи ИИ по основной причине – из-за отсутствия их автора или правообладателя как субъекта (не в категориях недоступного, неустановленного, а неизвестного). Между тем предусмотренные законом и применимые для установления автора сиротских произведений меры его поиска могли бы стать экспериментальной моделью идентификации правообладателей контента, который генерируется ИИ благодаря обучению на материалах им принадлежащих. В частности, такими мерами поиска могли бы стать обращение к открытым источникам информации, соответствующим данной категории объектов авторских или смежных прав; обращение в организации, управляющие правами на коллективной основе, или использование общедоступных информационных систем таких организаций (пункт 5 статьи 1243 ГК РФ); использование поисковых и информационных систем; поиск по библиотечным и архивным фондам. Правоприменительный аспект установления правообладателей контента, который генерируется ИИ при его обучении, мог бы предусматривать процедуры использования соответствующего объекта авторских и смежных прав путем обращения в аккредитованную организацию по коллективному управлению правами с заявлением о предоставлении пользователю (разработчику, владельцу сервиса) ИИ-контента права на использование соответствующего объекта за вознаграждение, подлежащее зачислению на специальный номинальный счет до установления правообладателя либо лица, уполномоченного на его получение. Для иллюстрации действительности количественного

масштаба сиротских произведений достаточно привести следующую статистику: с 2024 года Российское авторское общество, которое осуществляет управление авторскими правами на коллективной основе и которому делегированы полномочия на управление орфанными произведениями, зарегистрировало 900 из 28 000 таковых.

Определенная фантасмагоричность идеи о такой модели управления правами интеллектуальной собственности в отношении ИИ-контента может обосновываться вполне практическими реалиями. Прежде всего это позволит предупреждать риски ответственности, продуцированные заимствованием принадлежащих конкретным правообладателям объектов интеллектуальной собственности в генерируемом ИИ-контенте, который используется работниками для создания служебных произведений. Эта проблема детально освещалась в ранее изданном авторском исследовании [3], поэтому, избегая излишней тавтологии при самоцитировании, важно обратиться к новым эмпирическим данным.

Анализ данных ИИ-индустрии позволяет отметить проблему сиротских произведений в контексте сочетания экономических и правовых аспектов. Так, в мае 2026 года Microsoft отменил внутренние лицензии доступа работников к большим языковым моделям Anthropic (Claude AI) из-за перехода на оплату ИИ по токенам, взимающим плату за каждый сгенерированный запрос и строку кода, который в течение считанных месяцев съедает годовые бюджеты компании. Между тем от 70 до 90% токенов, используемых для обучения всех мировых ИИ-моделей составляют сиротские произведения [5], у значительного числа которых, как можно предположить, есть неустановленные правообладатели, не обращавшиеся за выплатой вознаграждений. В связи с этим понятна озабоченность ИИ-разработчиков соблюдением прав создателей и владельцев контента на надлежащее указание авторства и вознаграждение только при воспроизводстве объекта целиком, с отказом от таких обременений при условии анализа (чтения, изображения) объекта генеративными ИИ-моделями для перевода в общественное пользование.

Применению сиротских произведений при обучении нейросетей благоприятствует отсутствие режима охраноспособности ими же создаваемого контента. В силу нормы статьи 1281 ГК РФ срок действия исключительного права на произведение действует в течение всей жизни автора и 70 лет после его смерти. В отношении ИИ-контента указанную норму распространить невозможно ввиду отсутствия субъекта авторского права. В такой ситуации можно констатировать, что материал, произведенный генеративными моделями, переходит в общественное достояние сразу же после создания и находится в неконтролируемом обороте пользования в режиме «вечного окна», в том числе с риском обнаружения объекта авторского права в какой-либо части контента. При этом вопросы авторства и охраноспособности ИИ-контента отражаются на проблемах патентообразования ИИ-решений при высокой заинтересованности компаний в потенциале их коммерциализации. По данным Роспатента, за I квартал 2026 года патенты оформлены по 97 заявкам из 900 представленных.

Для предупреждения правовых и репутационных рисков использования ИИ-инструментов и одновременно для защиты интеллектуальной собственности от утраты прав на нее при выполнении служебных задач работниками юридических лиц вне зависимости от организационно-правовых форм, равно как и в собственной их операционной деятельности, следует отметить важность правового регулирования возникающих отношений. Целесообразно в трудовом договоре с работниками предусмотреть комплекс положений о распределении исключительных прав на служебные результаты интеллектуальной деятельности (РИД) и применении ИИ. Существенными для предупреждения указанных рисков являются следующие ключевые положения договора:

- РИД является служебным, если он создан в рамках должностных обязанностей или по заданию работодателя. РИД не признается служебным при создании за рамками обязанностей и заданий;

- работник несет ответственность за нарушение интересов правообладателей при выполнении должностных обязанностей (исполнении служебных полномочий) по созданию служебных произведений при помощи ИИ-сервисов;
- при осуществлении трудовых функций запрещено осуществлять передачу в публичные ИИ-системы информации, которая содержит коммерческую тайну, персональные данные, внутренние документы организации;
- работник обязан проверять и верифицировать результаты, используемые при создании служебных произведений, сгенерированные ИИ-сервисами.

Помимо этого, обязательными являются разработка и принятие локального нормативного акта, который определяет регламент использования ИИ-сервисов и правила работы в них и который также должен быть интегрирован с правилами внутреннего трудового распорядка, положением о коммерческой тайне, положением о защите служебной информации, политикой обработки персональных данных. Целью акта является обеспечение законного и безопасного использования ИИ-инструментов в деятельности организации, а обязательства его соблюдения должны распространяться на всех сотрудников, внедряющих и эксплуатирующих их. Целесообразно также нормативно определить область не только разрешенного, но и запрещенного применения ИИ-инструментов – в частности, исключающего нарушение пункта 6 части 1 статьи 86 ТК РФ, в соответствии с которой при принятии решений, затрагивающих интересы работника, работодатель не имеет права основываться на персональных данных, полученных исключительно в результате их автоматизированной обработки. Данное положение актуально в контексте использования ИИ-инструментов для проведения оценки персонала и обработки соответствующих данных. Обязательным при этом должно быть применение норм аудита и оценки рисков ИИ-систем (как внешних, так и корпоративных) в диагностике недопущения дискриминации, проверки безопасности, соблюдения законодательства и локальных нормативных актов, регулирующих вопросы защиты информации, персональных данных и соблюдения прав интеллектуальной собственности.

Рекомендации по управлению правовыми рисками и предупреждению неблагоприятных последствий в результате их воплощения в контексте смежных отношений, связанных с защитой прав интеллектуальной собственности и применением ИИ, распределяются по трем уровням назначения.

Прикладной уровень предполагает соблюдение следующих правил: наличие обязательного элемента промпта в запросе сведений о правообладателе материала в генерируемом контенте (ссылки, авторы, реквизиты издания); при публичном воспроизведении установление (в зависимости от обстоятельств) маркировок «создано с использованием ИИ», «создано без ИИ», «публичное воспроизведение материала, допускается с согласия правообладателя», «запрещено использование материала для интеллектуального анализа данных»; использование сервисов-детекторов ИИ в текстах, в том числе позволяющих выявлять маскировки синтетического материала («Думейт», «Дутрейс»).

Организационно-правовой уровень предусматривает разработку и принятие указанного выше нормативного акта с примерным названием «Об использовании ИИ», а также рекомендаций по недопущению этических дилемм при работе с ИИ при выполнении должностных обязанностей и при инициативном создании служебных РИД. Помимо этого, нельзя допускать упущения перспектив кадрового администрирования трудовых функций работников в позициях ИИ-комиссаров, а также промптеров в следующем наборе их квалификации: доменная экспертиза, кросс-функциональное взаимодействие, направленное на обеспечение перевода бизнес-задач в технические решения; навыки проведения фактчекинга и верификации, промпт-инжиниринг при уверенном осуществлении оптимизации запросов, проведении диагностики галлюцинаций,

формировании библиотеки промптов. К вопросам кадрового администрирования также следует отнести актуализацию трудовых договоров в части распределения прав на созданные служебные РИД и применения ИИ, а также организацию обучения по этим аспектам для формирования должной культуры в указанных сферах реализации правоотношений.

Стратегический уровень требует внедрения технологий обеспечения защиты корпоративных ИИ-моделей, которые создают барьер между внешней/корпоративной сетью и ИИ-сервисами. Такие технологии позволяют отслеживать и блокировать передачу данных (информации ограниченного распространения) или обрабатывать их без нарушения прав, с обеспечением надлежащей защиты и управлением рисками несоблюдения законодательства.

Предлагаемые рекомендации актуальны в свете публичного обсуждения законопроекта «Об основах государственного регулирования сфер применения технологий искусственного интеллекта в Российской Федерации», применение положений которого можно оценить в условиях детерминации представленных правовых репутационных рисков использования ИИ. В частности, к таким положениям можно отнести условия, разрешающие разработчикам ИИ использовать для обучения ИИ доступный или публично представленный контент (статья 5 законопроекта), что открывает простор для автоматизированного сбора данных при помощи специальных программ в условиях рисков нарушения интересов правообладателей. Благоприятные условия для нарушения прав интеллектуальной собственности несут проектируемые новеллы (статья 13), допускающие использование для обучения ИИ правомерно полученных экземпляров произведений, в том числе при доведении информации до всеобщего сведения или предоставлении ее для анализа (статья 13). Критичным в потенциале образования деликтов и нарушений в сфере интеллектуальной собственности является и ряд иных положений законопроекта. Учитывая, что природа их образования может стать нормативно обусловленной, крайне важно не допускать игнорирования представленных в настоящей статье предложений и рекомендаций по управлению ИИ-рисками в правовом содержании мер.

Список источников

1. Федеральный закон от 22.07.2024 № 190-ФЗ «О внесении изменений в часть четвертую Гражданского кодекса Российской Федерации».
2. Артений Л.С. Искусственный интеллект в авторском праве // Вестник науки и образования. 2019. № 7–1 (61).
3. Астанин В.В. О рисках и стандартах применения генеративного искусственного интеллекта при создании служебных РИД // Право интеллектуальной собственности. № 4. 2025. С. 25–28.
4. Балашова А.И. Искусственный интеллект в авторском и патентном праве: объекты, субъектный состав правоотношений, сроки правовой охраны // Журнал Суда по интеллектуальным правам. № 2 (36). Июнь 2022 года. С. 90–98.
5. Статистика Common Crawl (репозитарий данных веб-сканирования) (дата обращения: 25.05.2026).
6. Saudi Arabia bestows citizenship on a robot named Sophia (дата обращения: 20.05.2026).

Ознакомиться с презентацией

ОРГАНИЗАЦИЯ ФУНКЦИИ АУДИТА МОДЕЛЕЙ: 9 ЛЕТ ПРАКТИКИ

Служба внутреннего аудита рассматривает модель как элемент бизнес-процесса и проводит оценку не только ее функционирования, но и потенциального влияния ограничений модели на клиентский опыт, финансовый результат и уровень рисков.

О.В. ЧИСТЯКОВ
Руководитель службы внутреннего аудита ПАО Сбербанк

Аннотация

В статье представлена организация функции аудита моделей в Службе внутреннего аудита ПАО Сбербанк (далее – Служба), а также рассмотрены ключевые этапы ее становления в условиях расширения применения математических моделей в банковских процессах. Выделены основные элементы организации аудита моделей: формирование команды, развитие технологической инфраструктуры и использование механизмов сохранения и масштабирования экспертизы.

Ключевые слова: аудит моделей, модельный риск, валидация моделей, внутренний аудит, банковские процессы, система управления модельным риском.

Коды JEL: G21, G28, M42, C53.

Введение

Финансовый сектор претерпел масштабную трансформацию, обусловленную внедрением математических моделей в ключевые бизнес-процессы. Применение моделей позволило существенно повысить скорость и качество принятия решений. Вследствие этого модели, например, кредитного скоринга, оценки риска, антифрода, управления ликвидностью и клиентской аналитики стали неотъемлемой частью процесса принятия решений в банках.

Вместе с ростом проникновения моделей закономерно возрос и уровень модельного риска – риска принятия ошибочных решений вследствие недостатков, возникающих при проектировании, разработке, внедрении или эксплуатации моделей.

Значимость модельного риска обусловила развитие в банке системы независимого контроля за моделями. В рамках второй линии защиты в Сбере функционирует специализированное подразделение валидации, осуществляющее независимую проверку корректности разработки и функционирования моделей. В свою очередь, Служба (третья линия защиты) рассматривает модель как элемент бизнес-процесса и проводит оценку не только ее функционирования (в том числе с использованием дополнительных источников данных), но и потенциального влияния ограничений модели на клиентский опыт, финансовый результат и уровень рисков.

Необходимость развития функции аудита моделей в Службе была обусловлена стремительным ростом использования моделей в банковских процессах. Становление функции можно условно разделить на два этапа:

- до введения регуляторных требований;
- после их введения.

История организации функции аудита моделей в Службе внутреннего аудита ПАО Сбербанк

До введения регуляторных требований

В 2010 году ПАО Сбербанк внедрило первые модели в кредитный процесс. На начальном этапе подходы к проверкам Служба формировала преимущественно эмпирически, при этом накапливалась практика аудита моделей.

Служба проводила проверки отдельных наиболее значимых моделей, оказывающих существенное влияние на принятие решений в банке. Результаты таких проверок использовались при совершенствовании моделей, процессов их применения и снижении связанных с ними рисков. Практика проведения таких проверок была востребована банком и впоследствии стала основой для развития функции аудита моделей.

После введения регуляторных требований

В 2015 году Банк России утвердил требования к системе управления рисками[1], а также нормативные требования к применению моделей в рамках подхода внутренних рейтингов (ПВР)[2]. Развитие регулирования стало важным фактором институционализации модельного риска в качестве самостоятельного объекта контроля.

В 2017 году значительное увеличение количества моделей в банке, а также требование Банка России по ежегодной проверке всех существенных рисков потребовали реализации системного подхода к аудиту модельного риска. Вследствие этого в Службе было принято решение о создании отдела аудита моделей, в котором была консолидирована вся экспертиза, накопленная Службой по данному вопросу.

К 2018 году модели, участвующие в принятии решений, были внедрены в большинство процессов банка. Расширение масштабов их применения обусловило необходимость последующих организационных изменений в Службе. В связи с этим в 2020 году был создан распределенный хаб (территориально распределенная команда специалистов из 12 филиалов банка) в 11 часовых поясах по всей стране. Такая организация работы позволила оптимизировать нагрузку на ИТ-инфраструктуру банка при работе с массивами данных и увеличить продолжительность «операционного дня» Службы до 18 часов в сутки, а также обеспечить постоянный приток в Службу ИТ-кадров из регионов.

В настоящее время Служба на ежегодной основе проводит аудит как ПВР-моделей, находящихся под надзором регулятора, так и моделей, участвующих в принятии решений в конкретном проверяемом процессе. Таким образом, аудит моделей обеспечивает комплексную оценку системы управления модельным риском в целом.

---

1 Указание Банка России от 15.04.2015 № 3624-У «О требованиях к системе управления рисками и капиталом кредитной организации и банковской группы».

2 Положение Банка России от 06.08.2015 № 483-П «О порядке расчета величины кредитного риска на основе внутренних рейтингов».

Организационная функция аудита моделей: люди, технологии и фабрика знаний

Создание эффективной функции аудита моделей требует не только накопления экспертизы в части проведения проверок, но и выстраивания соответствующей организационной модели. Практика Службы показывает, что зрелость функции определяется сбалансированным развитием трех обязательных элементов:

- люди;
- технологии;
- знания (сохранение экспертизы).

Совокупное развитие данных компонентов позволяет обеспечить как необходимую глубину анализа моделей, так и устойчивость функции в долгосрочной перспективе.

Люди

Для проведения аудита моделей требуются две категории специалистов: эксперты в области аудита и специалисты по анализу данных и моделированию (Data Scientists).

Категорию экспертов в области аудита составляют сотрудники Службы со стажем работы более 5 лет, обладающие глубоким пониманием банковских процессов, структуры данных, а также знанием регуляторных требований. Кроме того, такие эксперты имеют навыки управления проектами и взаимодействия с владельцами процессов, что критически важно для проведения аудита моделей. Формирование экспертизы происходит в результате многолетнего накопления практического опыта в ходе проверок процессов.

Вторая категория специалистов – Data Scientists – должна обладать знаниями в области математической статистики, современных методов моделирования, а также навыками обработки больших данных и разработки моделей. Пополнение кадрового состава Службы такими специалистами осуществляется за счет внешнего рынка.

Сочетание указанных компетенций позволяет Службе одновременно оценивать корректность математической конструкции модели и соответствие результатов ее работы бизнес-целям, процессному контексту и требованиям управления рисками.

В результате аудит моделей выполняет не только контрольную, но и экспертную функцию, позволяя выявлять скрытые для владельцев процессов ограничения моделей и ошибки, приводящие к принятию некорректных решений и возникновению потенциальных рисков.

Технологии

Аудит моделей не может быть осуществлен исключительно на основе анализа документации или выборочных тестов, поскольку в ходе проверки требуется подтверждение корректности результатов моделей, их переобучение, тестирование устойчивости и анализ альтернативных гипотез. Для решения этих задач в банке сформирована технологическая инфраструктура, обеспечивающая доступ Службы к значительным вычислительным мощностям и хранилищу данных, а также к широкому набору программных платформ и инструментов автоматизации.

Вычислительная среда, используемая Службой, включает:

- более 300 CPU (Central Processing Unit – центральное процессное устройство) для обработки данных;

- около 50 GPU (Graphics Processing Unit – графический процессор) для выполнения вычислительно емких операций, связанных с переобучением, тестированием и исследованием моделей;
- около 45 терабайт хранилища данных закреплено непосредственно за Службой для аудита моделей, что позволяет работать с историческими массивами информации и большими объемами данных, необходимыми для оценки устойчивости моделей и воспроизводимости результатов.

Дополнительные возможности обеспечиваются широким набором аналитических платформ и инструментов. Используемый технологический стек включает Python, инструменты Process mining, Code mining и другие решения, позволяющие проводить углубленный анализ процессов, исходного кода моделей и корректности их функционирования.

Наряду с применением стандартных аналитических платформ, в Службе разработано собственное программное решение для автоматизации валидационных тестов. Данный инструмент позволяет существенно сократить трудоемкость проверок, повысить воспроизводимость результатов и минимизировать риск человеческой ошибки.

Таким образом, для реализации функции аудита моделей необходимы технологическая инфраструктура, значительные вычислительные мощности и большие данные.

Фабрика знаний

Третьим необходимым элементом для развития функции аудита моделей является система накопления и передачи знаний. Для высокотехнологичных направлений аудита традиционно характерна зависимость от специалистов – носителей экспертизы. Уход таких специалистов может приводить к утрате критически важных знаний, снижению качества проверок и дополнительным затратам, связанным с поиском новых сотрудников и их адаптацией.

Как и для других высокотехнологичных направлений, для функции аудита моделей характерен риск потери экспертизы вследствие высокой мобильности ИТ-специалистов. В этих условиях одной из ключевых задач становится сохранение и передача накопленных знаний. В Службе данная проблема решается с помощью внедрения фабрики знаний – инструмента сохранения экспертизы и развития компетенций.

Основу данного инструмента составляет онтология, включающая более 40 предметных областей, что обеспечивает структурированную навигацию по знаниям, единообразие терминологии и формирование семантических связей между различными направлениями экспертизы.

Данные фиксируются в виде карточек знаний в человеко- и машиночитаемом формате, что позволяет не только сохранять результаты проверок (извлечение знаний), но и использовать их для интеллектуального поиска с целью повторного применения наработанных практик. Обязательным условием развития фабрики знаний является применение генеративного ИИ для формирования ответов (загрузка знаний) на запросы сотрудников/новичков на основе накопленной базы знаний.

Принципиально важно, что пополнение фабрики знаний является обязательным этапом завершения каждой аудиторской проверки. Это позволяет накапливать экспертизу и, как следствие, снижать зависимость функции от конкретных специалистов, обеспечивая передачу накопленного опыта новым сотрудникам.

Выводы

По мере расширения применения моделей в банковских процессах трансформируется и роль внутреннего аудита – от проверки корректности отдельных моделей к комплексной оценке системы управления модельным риском. В этих условиях ключевым фактором эффективности становится способность внутреннего аудита обеспечивать глубокое понимание ограничений моделей, идентификацию источников модельного риска и оценку потенциального влияния ошибок моделирования.

На основе многолетней практики Службы мы видим, что только комплексный подход, включающий формирование кросс-функциональной команды, создание и эффективное использование технологической инфраструктуры, внедрение механизма сохранения и масштабирования экспертизы, позволяет полноценно реализовать функцию аудита моделей.

Список источников

1. Положение Банка России от 08.04.2020 № 716-П «О требованиях к системе управления операционным риском в кредитной организации и банковской группе».
2. Положение Банка России от 27.11.2024 № 845-П «О порядке расчета кредитного риска кредитными организациями на основе внутренних рейтингов».
3. Указание Банка России от 15.04.2015 № 3624-У «О требованиях к системе управления рисками и капиталом кредитной организации и банковской группы».
4. Basel Committee on Banking Supervision. International Convergence of Capital Measurement and Capital Standards: A Revised Framework (Comprehensive Version). Basel: Bank for International Settlements, 2006.

Ознакомиться с презентацией

АУДИТ ЖИЗНЕННОГО ЦИКЛА МОДЕЛЕЙ: НОВЫЕ ВЫЗОВЫ И РАЗВИТИЕ ПОДХОДОВ

Аудит жизненного цикла моделей постепенно превращается из проверки отдельных элементов в оценку устойчивости всей системы управления модельным риском. Нас интересует уже не только качество алгоритма. Нас интересует качество среды, в которой этот алгоритм живет, развивается и принимает решения.

М.А. МАМАЛАЕВА
Руководитель дирекции внутреннего аудита
АО «Альфа-Банк»

Аннотация

В статье рассматривается трансформация подходов к аудиту жизненного цикла моделей в условиях перехода от статистических моделей к генеративному ИИ. Анализируются ключевые тенденции: расширение понятия модельного риска, переход от периодической проверки к непрерывному мониторингу, проблема незаявленных моделей вне официального реестра и необходимость перехода на специализированные платформы управления жизненным циклом моделей. Обосновывается тезис об эволюции аудита моделей в аудит систем принятия решений организации в целом, требующий от аудитора понимания не только логики алгоритма, но и экосистемы вокруг него.

Ключевые слова: модельный риск, жизненный цикл модели, генеративный искусственный интеллект, внутренний аудит, непрерывный мониторинг, незаявленные модели, платформы управления моделями, ИИ-агенты.

Коды JEL: G21, G32, M42, O33.

Почему 2025 год стал годом переосмысления системы управления модельными рисками?

Долгое время выстроенная система управления модельными рисками давала нам структуру и уверенность. Она была нашим надежным компасом в мире статистических моделей. Мы понимали правила игры. На вход поступали данные, на выходе появлялся результат, а между ними была понятная логика. Да, модели ошибались и порой требовали донастройки. Однако в целом они были предсказуемыми.

Но мир меняется. Сегодня рядом с нами работают генеративные модели, глубокие нейронные сети, адаптивные алгоритмы и системы ИИ, которые уже не вписываются в привычные рамки. Они учатся. Меняются. Переобучаются. Дрейфуют. Иногда складывается ощущение, что модель живет собственной жизнью: вчера она принимала решения одним способом, а сегодня – уже немного другим. И это не ошибка. Скорее – новая реальность. Темпы внедрения новых технологий оказались настолько высокими, что наши процессы управления не всегда за ними успевают. Пока согласовывается одна методика, появляется уже следующая версия модели.

Но дело не только в скорости. Изменился сам периметр принятия решений. Когда-то модели помогали оценивать кредитные риски. Сегодня они участвуют в управлении активами и пассивами, помогают принимать решения по капиталу, прогнозируют поведение клиентов, выявляют мошенничество и становятся частью стратегических процессов организации. При этом логика их работы зачастую напоминает черный ящик. Мы видим результат, но не всегда можем легко

объяснить путь, который привел к этому результату. И проблема не в том, что старые подходы перестали работать. Они по-прежнему важны и полезны. Просто сегодня их уже недостаточно. Мир стал сложнее, а значит, и наши инструменты должны стать более глубокими и гибкими.

Так что же нам, как внутреннему аудиту, остается делать? Одно дело – понимать, что объекты аудита становятся все более сложными и непрозрачными. Совсем другое – превратить это понимание в ежедневную практику.

Хорошая новость заключается в том, что мы не начинаем этот путь с нуля. Когда-то все дебютировало с аудитов качества рисковых моделей в кредитовании физических и юридических лиц (ПВР). Затем появились более сложные задачи. Сегодня мы рассматриваем жизненный цикл модели целиком: от ее появления и постановки на учет до валидации, промышленного внедрения, мониторинга качества, модернизации и последующего вывода из эксплуатации.

Но по мере того как модели становятся все более динамичными, меняется и сам подход к аудиту. Если раньше мы проверяли относительно статичный объект, то сегодня работаем с системой, которая способна меняться практически в режиме реального времени. Именно поэтому аудит жизненного цикла моделей постепенно превращается из проверки отдельных элементов в оценку устойчивости всей системы управления модельным риском. Нас интересует уже не только качество алгоритма. Нас интересует качество среды, в которой этот алгоритм живет, развивается и принимает решения.

Какие тенденции будут определять развитие управления модельными рисками в ближайшие годы? Прежде всего мы наблюдаем расширение самого понятия модельного риска. Сегодня перечень направлений, где применяются модели, растет практически ежемесячно. Модели используются в антифрод-системах, клиентской аналитике, AML-процессах, ценообразовании, прогнозировании и во множестве других направлений. Каждая такая модель способна влиять на бизнес-решения. Каждая может создавать риски.

Вторая тенденция связана с непрерывным мониторингом. Современный бизнес хочет видеть состояние модели здесь и сейчас. По сути, происходит переход от периодических «медицинских осмотров» к постоянному мониторингу жизненно важных показателей. И это вполне логично. Никто не хочет узнавать о проблеме тогда, когда она уже успела перерасти в катастрофу.

Третья тенденция, пожалуй, самая любопытная. Все большее значение приобретают так называемые незаявленные модели – от простых до сложных, которые могут участвовать в принятии решений, оставаясь вне официального реестра. И здесь возникает простой, но очень важный вопрос: а все ли модели в организации мы видим?

Наконец, становится очевидным, что ручное управление модельными рисками перестает масштабироваться. Когда моделей несколько десятков, еще можно полагаться на таблицы, реестры и отдельные документы. Когда их сотни или тысячи, такой подход начинает напоминать попытку управлять современным аэропортом с помощью бумажного ежедневника. Именно поэтому организации все активнее переходят к специализированным платформам управления жизненным циклом моделей. Они позволяют объединить процессы учета, валидации, мониторинга, управления изменениями и отчетности в единую систему.

В 2026 году аудит управления модельными рисками – это уже не только аудит моделей. Это аудит систем принятия решений организации. И если раньше аудитору достаточно было понимать математику модели, то сегодня ему все чаще приходится разбираться в экосистеме вокруг нее.

Возможно, через несколько лет модели будут принимать решения быстрее, чем люди успеют их проанализировать. Возможно, часть процессов станет полностью автономной.

Возможно, аудиторам придется проверять уже не отдельные модели, а целые сообщества взаимодействующих интеллектуальных агентов.

Но одна вещь останется неизменной. Внутренний аудит должен быть полезной и успешной функцией. Поэтому очень важно то, как мы справимся с новым вызовом, которым является аудит системы управления модельными рисками, в том числе содержащими AI-агентов и генеративные модели, где нам есть куда расти.

Список источников

1. Five Model Risk Management Trends Defining 2026. 10.02.2026, YIELDS.
2. Why Model Risk Management Needs to Be a 2026 Priority By Andrew Pery, AI Ethics Evangelist, ABBYY. 21 November 2025, AI Journal.

Ознакомиться с презентацией

УПРАВЛЕНИЕ РИСКАМИ НОВЫХ ТЕХНОЛОГИЙ: ОПЫТ МОСКОВСКОЙ БИРЖИ

Московская Биржа ориентируется на работу по принципу «человек в контуре». Этот подход обеспечивает более высокий уровень управляемости, минимизирует ошибки при работе ИИ и подчеркивает ответственность пользователя за финальное решение.

С.В. ДЕМИДОВ
Заместитель председателя правления по информационной безопасности Группы «Московская Биржа»

Аннотация

В статье представлен практический опыт Московской Биржи по интеграции управления ИИ-рисками в существующую систему риск-менеджмента. Описан разработанный фреймворк, базирующийся на принципе «человек в контуре» и методологии MLSecOps, охватывающий полный жизненный цикл сложных ИИ-систем (генеративных, рекомендательных и предиктивных). Детально раскрывается критериальная модель оценки из 12 параметров, позволяющая количественно определять уровень риска через «куб риска» и визуализировать его на тепловых картах. Особое внимание уделено адаптации классических механизмов управления операционными рисками под специфику алгоритмических систем и обеспечению безопасности на этапах планирования, разработки и эксплуатации.

Ключевые слова: управление ИИ-рисками, фреймворк рисков, принцип «человек в контуре», MLSecOps, оценка рисков, жизненный цикл ИИ, «куб риска».

Коды JEL: G20, O33, M15, D81

В условиях стремительной цифровизации и усложнения алгоритмических систем критически важным является управление рисками, связанными с новыми технологиями. Особое место среди них занимает ИИ, который становится самостоятельным источником риска, находящимся на стыке операционных, кибер-, этических и комплаенс-рисков. Московская Биржа разработала комплексный подход (фреймворк) к управлению такими рисками, а также успешно интегрировала его в существующую систему риск-менеджмента, чтобы обеспечить внедрение инноваций без снижения риск-защищенности.

Основная используемая терминология соответствует Указу Президента Российской Федерации № 490[1], в котором ИИ-модель признается программой электронных вычислительных машин, выполняющих задачи на одном уровне с человеком, а ИИ-система – технической системой, рождающей прогнозы и рекомендации для решения заданного набора человеческих задач.

Ключевым вызовом использования ИИ в деятельности компании является определение роли человека в процессе принятия решений, основанных на рекомендациях ИИ-систем. Возможны два уровня управляемости:

1. «Человек в курсе»: ИИ-модель самостоятельно принимает решение, а человек лишь осведомлен об этом результате или проверяет его постфактум.

[1] Указ Президента Российской Федерации от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации».

2. «Человек в контуре»: человек сохраняет решающую роль, принимая финальное решение на основе рекомендаций, предоставленных ИИ-моделью.

Московская Биржа ориентируется на работу по принципу «человек в контуре». Этот подход обеспечивает более высокий уровень управляемости, минимизирует ошибки при работе ИИ и подчеркивает ответственность пользователя за финальное решение.

Управление ИИ-рисками сфокусировано на сложной и нелинейной природе трех типов ИИ: генеративном, рекомендательном и предиктивном. Риски, связанные с более простыми алгоритмами, такими как линейные нейросети, продолжают управляться в рамках стандартных процедур.

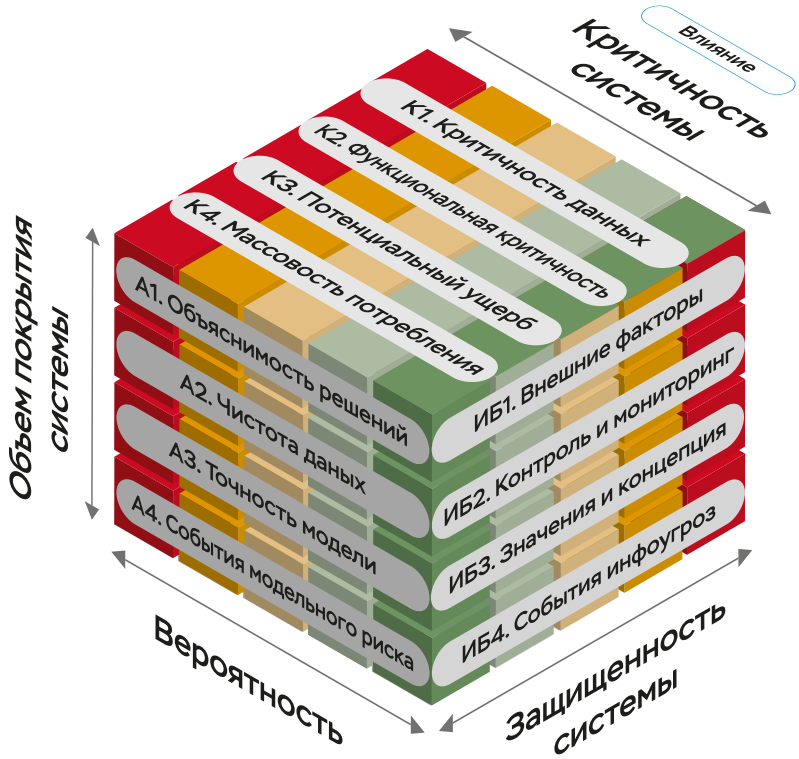
ИИ-риски рассматриваются как подвид нефинансовых рисков. При этом распределение ролей, принципы систем управления и ответственности линий защиты не являются уникальными, а эффективно интегрируются в уже существующую систему управления рисками (СУР) в рамках классических трех линий защиты. Это позволяет использовать уже отработанные механизмы, адаптируя их под специфику ИИ.

Управление рисками осуществляется на всех этапах жизненного цикла ИИ-системы. Для стадий разработки, тестирования и развертывания используется продуктовый риск-менеджмент, в котором происходит проактивное управление рисками будущей ИИ-системы и заранее реализуются соответствующие меры, выстраивается система контролей, чтобы минимизировать будущие риски.

На этапе эксплуатации используется классический риск-менеджмент, который не только актуализирует информацию о выявленных ранее или новых идентифицированных рисках, но и покрывает работу с рисковыми событиями.

Для оценки риска по конкретной ИИ-системе разработан критериальный подход, основанный на двух фундаментальных параметрах: значимости и вероятности риска. Как показано на рисунке, оценка производится по 12 критериям, каждый из которых оценивается по пятибалльной шкале.

«КУБ РИСКА» ДЛЯ ОЦЕНКИ ИИ-СИСТЕМЫ



1. Значимость риска (критичность системы):
   - критичность данных: категория данных, используемых при разработке (конфиденциальность, персональные данные);
   - функциональная критичность: сфера применения и критичность бизнес-процессов, для которых используется ИИ;
   - потенциальный ущерб: размер возможных убытков, регуляторных штрафов или репутационного ущерба;
   - массовость потребления: количество клиентов или внутренних пользователей, применяющих систему.

2. Вероятность риска.

   Оценивается через два блока:
   2.1. ИИ-система как модель:
   - объяснимость решений: наличие механизмов валидации и понимания логики ИИ;
   - чистота данных: качество и происхождение обучающих наборов данных;
   - точность модели: технико-статистические параметры (Accuracy, Error Rate и другие);
   - история риск-событий: наличие инцидентов, связанных с модельным риском.

   2.2. ИИ-система как ИТ-система:
   - зависимость от внешних факторов: участие третьих лиц, технологические ограничения;
   - контроль и мониторинг: наличие механизмов защиты от атак, системной безопасности;
   - компетенции: уровень знаний владельцев, разработчиков и пользователей;
   - история инцидентов ИБ: наличие угроз информационной безопасности в отношении модели.

Результат оценки визуализируется в виде «куба риска», оценка вероятности и влияния позволяет определить совокупный уровень риска по ИИ-системе и принять решение о ее допуске к эксплуатации, о необходимости реализации дополнительных мер и контролей. В будущем планируется калибровать веса критериев в зависимости от типа ИИ.

Эффективное управление ИИ-рисками требует постоянного сбора данных о рисках / рисковых событиях. Для решения этой задачи предлагается использовать внутренние и внешние источники:
- внутренние: мониторинг ИТ-систем, анализ качества данных и уязвимостей ИБ, результаты валидации и аудита модели, данные по уязвимостям и классические процедуры управления рисками;
- внешние: нормативные акты Российской Федерации и международные стандарты, отраслевые исследования, открытые базы знаний, экспертиза профессиональных сообществ и другие.

Таким образом, опыт Московской Биржи демонстрирует возможность построения эффективной системы управления ИИ-рисками на базе существующей инфраструктуры риск-менеджмента. Интеграция принципов MLSecOps, принцип «человек в контуре» и детальная критериальная оценка позволяют безопасно внедрять сложные современные технологии в деятельность компании. Фреймворк по управления ИИ-рисками находится в стадии непрерывного развития и совершенствования, адаптируясь к быстро меняющемуся технологическому ландшафту и новым регуляторным требованиям.

Список источников

1. Доклад для общественных консультаций «Применение искусственного интеллекта на финансовом рынке: текущий статус и условия дальнейшего развития» (дата обращения: 15.07.2026).
2. Кодекс этики в сфере разработки и применения искусственного интеллекта на финансовом рынке (дата обращения: 15.07.2026).
3. Методические рекомендации Банка России от 16.06.2026 № 3-МР по обеспечению информационной безопасности при разработке и применении искусственного интеллекта на финансовом рынке (дата обращения: 15.07.2026).
4. AI Secure Agentic. Framework Essentials (AI-SAFE) v 1 (дата обращения: 15.07.2026).
5. Google's Secure AI Framework: A practitioner's guide to navigating AI security (дата обращения: 15.07.2026).
6. OWASP Top-10 для LLM и генеративного ИИ (2025) (дата обращения: 15.07.2026).

Ознакомиться с презентацией

РИСКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ВЫЗОВЫ ДЛЯ ВНУТРЕННЕГО АУДИТА И ВНУТРЕННЕГО КОНТРОЛЯ

Если ранее ключевое значение имели преимущественно межличностные формы взаимодействия в рамках бизнес-процессов, то в условиях интеграции ИИ-решений возникает необходимость формирования новых компетенций, включая постановку задач для ИИ-систем, интерпретацию и верификацию полученных результатов, а также развитие критического мышления.

А.В. ДАВЫДОВА
Руководитель группы внутреннего аудита, анализа и контроля рисков ООО «Технологии доверия – консультирование»

Аннотация

В условиях стремительной цифровизации ИИ перестает быть просто технологическим трендом и превращается в ключевой инструмент управления. Однако по статистике значительная часть ИИ-проектов не оправдывает изначальных ожиданий. В статье разбираются системные причины таких провалов, на которые специалистам в области внутреннего контроля и внутреннего аудита следует обратить внимание в своей работе. Отдельно рассматривается вопрос ответственности за работу ИИ-инструментов, а также вопросы развития компетенций сотрудников. В публикации представлена адаптированная модель цифрового доверия, которую внутренние аудиторы могут взять за основу проведения анализа ИИ-решений с точки зрения различных аспектов.

Ключевые слова: искусственный интеллект, информационные технологии, цифровая трансформация, риски, модель цифрового доверия.

Коды JEL: M42, M15, O33, D81.

В настоящее время наблюдается последовательное и устойчивое расширение практики использования решений на базе ИИ в операционной деятельности организаций. В силу технологической специфики данных решений существенное внимание традиционно уделяется техническим и эксплуатационным рискам, включая непрозрачность алгоритмов (эффект черного ящика), ограниченную интерпретируемость результатов, а также склонность моделей к генерации некорректных или недостоверных выводов (галлюцинациям).

Вместе с тем фокусировка исключительно на данных рисках является недостаточной. Существует широкий спектр сопутствующих рисков, которые требуют должного внимания как со стороны руководства организаций, так и со стороны внутреннего аудита и внутреннего контроля.

Инвестиционный цикл и искусственный интеллект как объекты инвестирования

Инвестиции организаций в решения на базе ИИ демонстрируют устойчивый рост. По оценкам Gartner, в 2026 году ожидается увеличение инвестиций в ИИ-технологии на 44%. При этом значительная часть инициатив не достигает запланированных результатов. Ключевыми причинами, снижающими вероятность успешной реализации проектов, являются:

- отсутствие четко определенной и измеримой бизнес-ценности внедрения ИИ-решений;
- недостаточная зрелость данных или их низкое качество;

- недооценка совокупной стоимости владения ИИ-инструментами;
- недостаточное внимание к управлению организационными изменениями.

При этом сама технология, лежащая в основе ИИ-решений, сравнительно редко выступает первопричиной неудачи проектов. Рассмотрим более подробно перечисленные ключевые причины.

Отсутствие четко определенной и измеримой бизнес-ценности внедрения ИИ-решений

Стремление к масштабированию эффектных демонстрационных кейсов и внедрению ИИ-решений во все функциональные области приводит к распылению ресурсов и снижению общей эффективности инвестиций. При отсутствии четко определенных критериев приоритизации и показателей эффективности такие инициативы не формируют устойчивой ценности и становятся уязвимыми в условиях бюджетных ограничений и повышенных требований к обоснованию рентабельности инвестиций.

В связи с этим во время оценки и отбора инвестиций в ИИ-решения критически важно формализовать целевые бизнес-результаты и обеспечить их последующий мониторинг на основе измеримых показателей – например, роста производительности, снижения операционных затрат и повышения удовлетворенности клиентов.

Недостаточная зрелость данных или их низкое качество

Качество данных является ключевым фактором успешного обучения и функционирования ИИ-решений. Несмотря на усиление внимания к вопросам управления данными в информационных системах в последние годы, именно недостаточность полноты, точности и согласованности данных нередко выступает существенным ограничением при внедрении и обучении ИИ-моделей.

При наличии одной и той же информации в нескольких информационных системах с разными, зачастую противоречащими друг другу значениями, ИИ не способен однозначно определить корректный вариант. Противоречия, которые должны устраняться на уровне управления данными, фактически становятся задачей для ИИ. В результате модели вынуждены опираться на все доступные источники при обучении и анализе, что приводит к искажению результатов и снижению их достоверности.

Во многих организациях отсутствуют формализованные подходы и метрики оценки качества данных. В таких условиях проблемы выявляются постфактум – на этапе получения некорректных результатов работы ИИ, что является наиболее затратным сценарием обнаружения ошибок.

Описанные проблемы качества данных не являются новыми. Однако использование ИИ существенно усиливает их влияние: ошибки и несогласованности данных начинают напрямую воздействовать на результаты работы, масштабируя и ускоряя распространение искажений.

Следует отметить, что в организациях финансового сектора исторически сформировались более зрелые практики управления данными и контроля их качества по сравнению с компаниями нефинансового сектора, что обусловлено более жесткими регуляторными требованиями и высокой значимостью данных для принятия решений.

Недооценка совокупной стоимости владения ИИ-инструментами

Рост затрат на поддержку ИИ-решений может привести к закрытию проектов даже в случаях их технической успешности и востребованности со стороны пользователей. Например, незначительная стоимость одного токена при масштабировании – с учетом тысяч пользователей и множества сценариев использования – трансформируется в существенные совокупные затраты владения.

Организации нередко недооценивают операционные расходы на поддержку ИИ-решений, поскольку на начальных этапах отсутствует полное понимание механизмов и темпов масштабирования затрат. В результате проекты, демонстрирующие жизнеспособность на стадии проверки концепции, на этапе промышленной эксплуатации могут становиться экономически неэффективными, что приводит к их пересмотру или закрытию.

Недостаточное внимание к управлению организационными изменениями

Без системного управления изменениями даже технически совершенные ИИ-инструменты демонстрируют низкий уровень внедрения и последующего использования. Со временем интенсивность их применения снижается, поскольку сотрудники воспринимают такие решения не как средство расширения профессиональных возможностей, а как источник угрозы. В результате организация реализует лишь незначительную долю потенциального эффекта от внедрения.

Ответственность за ошибки в работе ИИ-технологий

Руководители всех уровней – от глав бизнес-подразделений до топ-менеджмента – ежедневно вовлечены в многочисленные управленческие процессы. При этом с повышением уровня управления возрастает как объем, так и разнообразие обрабатываемой информации. В сложившейся практике принятия решений руководители во многом опираются на результаты работы и суждения подчиненных сотрудников. Важно понимать, что интеграция ИИ-технологий в бизнес-процессы не изменяет фундаментального принципа персональной ответственности: ответственность за принимаемые решения, включая юридически значимые последствия, сохраняется за человеком. Переложить ответственность на ИИ-решение не представляется возможным.

В этих условиях особую значимость приобретает выстраивание системы внутреннего контроля, обеспечивающей обоснованную уверенность в корректности функционирования ИИ-решений. Такая система должна предусматривать процедуры верификации результатов, выявления отклонений и потенциальных сбоев (включая некорректные или недостоверные выводы), а также четкое определение триггеров реагирования и распределение ответственности за принятие корректирующих мер.

Наем, обучение и развитие персонала в эпоху ИИ

Внедрение ИИ-технологий, как и любое организационное изменение, трансформирует карту компетенций персонала. Требуется целенаправленное развитие навыков эффективного взаимодействия с ИИ-инструментами. Если ранее ключевое значение имели преимущественно межличностные формы взаимодействия в рамках бизнес-процессов, то в условиях интеграции ИИ-решений возникает необходимость формирования новых компетенций, включая постановку задач для ИИ-систем, интерпретацию и верификацию полученных результатов, а также развитие критического мышления.

ИИ-технологии частично автоматизируют функции, традиционно выполняемые сотрудниками младших должностных уровней, что трансформирует роль входных позиций в организации. В этих условиях привычные модели профессионального становления и накопления практического опыта могут утрачивать эффективность. Это требует пересмотра подходов к найму и обучению молодых специалистов в среднесрочной перспективе (3–7 лет), включая смещение акцента с выполнения рутинных операций на развитие аналитических, интерпретационных и управленческих компетенций, а также навыков взаимодействия с ИИ-системами.

Модель цифрового доверия

С учетом рассмотренных аспектов влияния внедрения ИИ-инструментов на деятельность организаций особую значимость для специалистов по внутреннему контролю и внутреннему аудиту приобретает проблема обеспечения доверия к результатам их функционирования.

Всемирный экономический форум в рамках своей инициативы по цифровому доверию предложил модель цифрового доверия, разработанную в ответ на возрастающую потребность в обеспечении доверия к новым технологическим решениям и их взаимодействии с организациями и обществом. Указанная модель может служить концептуальной основой для анализа доверия к ИИ-системам с различных методологических позиций. Ниже рассматриваются подходы к адаптации данной модели для целей внутреннего контроля и внутреннего аудита в контексте анализа внедрения ИИ-технологий.

Модель включает три уровня, на каждом из которых определены области анализа, позволяющие оценивать доверие к системе внутреннего контроля в отношении ИИ-технологий в разрезе различных аспектов.

МОДЕЛЬ ЦИФРОВОГО ДОВЕРИЯ

1. Безопасность и надежность

- Конфиденциальность – область анализа, направленная на обеспечение надлежащего контроля за обработкой персональных данных и защиту информации организации от несанкционированного доступа, использования и раскрытия в рамках функционирования ИИ-систем.
- Кибербезопасность – область анализа, ориентированная на оценку защищенности ИИ-систем, включая используемые данные, технологии и связанные с ними процессы. Эффективная система внутреннего контроля рисков информационной безопасности ИИ-технологий обеспечивает снижение риска несанкционированного доступа и потенциального ущерба цифровым процессам и системам, а также обеспечивает их устойчивость и соблюдение принципов конфиденциальности, целостности и доступности данных и систем.
- Безопасность – область анализа, направленная на предотвращение ущерба, связанного с использованием технологий и обработкой данных. В контексте ИИ-систем особое значение приобретают аспекты не только информационной, но и физической безопасности. В частности, при эксплуатации беспилотных систем критическую роль играет обеспечение физической безопасности.

2. Подотчетность и надзор

Прозрачность – область анализа, направленная на обеспечение ясности и однозначности понимания интеграции ИИ-технологий в бизнес-процессы, включая используемые входные данные, логику функционирования и требования к действиям сотрудников. Обеспечение прозрачности способствует интерпретируемости результатов работы ИИ-систем и повышает их понятность для пользователей.

Возможность возмещения ущерба – область анализа, орентированная на оценку того, в какой степени риски, связанные с использованием ИИ-технологий, идентифицированы, оценены и покрыты механизмами внутреннего контроля. Ошибки и неопределенности, присущие функционированию ИИ-систем, могут приводить к возникновению непредвиденного ущерба. С учетом ограниченной практики передачи таких рисков (например, посредством страхования) особое значение приобретает определение приемлемого уровня риска и реализация мер по его снижению. В рамках данной области анализируются достаточность и эффективность контрольных процедур, направленных на минимизацию потенциального ущерба.

Проверяемость – область анализа, характеризующая способность организации, включая функцию внутреннего аудита, осуществлять независимую проверку и подтверждение корректности функционирования ИИ-систем, а также эффективности выстроенной системы внутреннего контроля.

3. Этичное и ответственное использование

Справедливость – область анализа, направленная на оценку того, приводит ли использование ИИ-технологий к дифференцированному воздействию на различные группы (например, отдельные категории клиентов), а также на анализ обоснованности и сопоставимости результатов для заинтересованных сторон. В данном контексте цели, связанные со справедливостью, охватывают как процедурные аспекты, так и характеристики конечных результатов. Процедурные аспекты включают оценку цифровых продуктов и услуг на предмет справедливости, их регулярный пересмотр, а также документирование принятых решений и применяемых подходов.

Совместимость – область анализа, характеризующая способность ИИ-систем взаимодействовать и обмениваться информацией для совместного использования без избыточных ограничений и барьеров. Особое значение совместимость приобретает в контексте развития мультиагентных ИИ-систем. Цели обеспечения совместимости сосредоточены преимущественно на технических аспектах, включая совместимость данных, платформ и интерфейсов, а также на согласованности архитектурных решений в рамках системы внутреннего контроля.

Список литературы

1. World Economic Forum, in collaboration with Accenture. Measuring Digital Trust: Supporting Decision-Making for Trustworthy Technologies. October 2023 (дата обращения: 15.07.2026).

Ознакомиться с презентацией

Значительно сокращается время между обнаружением проблемы, разработкой решения и оценкой эффекта от этого решения. В результате процессы и аудит выполняются синхронно, а решения принимаются здесь и сейчас на основании самых актуальных данных.

Г.Ю. ДЕЛЬВИГ
Начальник департамента внутреннего аудита
ПАО «Газпром нефть»

АУДИТ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ВЫЗОВ БУДУЩЕГО

Аннотация

В статье рассматривается трансформация функции внутреннего аудита в эпоху внедрения ИИ-технологий. Автор анализирует переход от традиционных методов проверки к использованию ИИ-агентов и продвинутых инструментов анализа данных. Особое внимание уделяется необходимости расширения цифровых компетенций аудиторов, созданию безопасной среды для проверки новых моделей и алгоритмов, а также формированию экосистемы, отвечающей потребностям компании и способствующей повышению эффективности бизнес-процессов.

Ключевые слова: искусственный интеллект в аудите, ИИ-агенты, аналитика больших данных, расширение компетенций, эффективность бизнес-процессов, «песочница» для тестирования, предиктивный аудит.

Коды JEL: M42, O33, L86.

Одним из главных вызовов для бизнеса в эпоху цифровизации является необходимость обработки беспрецедентного объема данных, который генерируют информационные системы. Как следствие, меняются стандарты производственных процессов и системы внутреннего контроля. Это в свою очередь напрямую влияет на функцию внутреннего аудита и его эффективность. Традиционные подходы к аудиту, основанные на выборочных проверках и ретроспективном анализе, не соответствуют актуальным технологическим требованиям. Параллельно с цифровой трансформацией бизнеса меняется и внутренний аудит, активно внедряя цифровые инструменты и ИИ в свою деятельность, при этом эволюционируя от наблюдателя и контролера до технологичного надежного советника, защищающего ценности компании.

Традиционная модель аудита постепенно уступает место непрерывному аудиту, что позволяет не просто реагировать на возникающие проблемы и риски, а своевременно предотвращать их. На помощь аудитору приходят различные цифровые инструменты. Для их эффективного использования требуются новые компетенции. В результате современный аудитор становится кросс-функциональным специалистом, который помогает бизнесу в освоении технологий, в выборе наиболее эффективных инструментов и подходов.

С каждым новым витком развития технологии позволяют анализировать показатели и процессы компании с большей точностью и детализацией. Буквально 2–3 года назад майнинг-процесс казался наилучшим решением для понимания фактического исполнения процессов и поиска проблемных зон для оптимизации. У широкого круга специалистов появилась возможность заглянуть в черный ящик корпоративных баз данных и в реальности оценить эффективность бизнес-процессов. Но этого уже недостаточно.

Сегодня развитие невозможно без внедрения ИИ и решений на его основе. ИИ становится неотъемлемой частью повседневной работы во всех сферах, и внутренний аудит не должен быть исключением.

С помощью ИИ аудитор может оценить всю совокупность данных, не ограничиваясь лишь выборкой. Классический цикл аудиторской проверки трансформируется в ответ на постоянно изменяющиеся факторы бизнес-среды:

- **Увеличение объема и сложности данных.** Ручной анализ невозможно применить к терабайтам ежедневно генерируемых данных, среди которых могут быть критически важные. ИИ-инструменты позволят проанализировать весь объем данных и выделить аномалии и отклонения.
- **Увеличение скорости принятия решений.** Бизнес-процессы постоянно автоматизируются и совершенствуются. Рекомендации, основанные на данных полугодовой давности, уже могут быть неактуальны. ИИ-агенты способны мониторить поток данных в реальном времени, помогая аудитору оперативно реагировать на инциденты или даже предотвращать их.
- **Скрытые (неочевидные) закономерности.** За счет оперативной обработки большого объема данных, ИИ может выявить сложные паттерны и корреляции, схемы мошенничества, а также неочевидные зоны улучшения и снижения операционных расходов.

Цикл оптимизации бизнес-процессов, таким образом, становится короче и эффективнее. Аудитор получает инструменты, позволяющие не только наблюдать за выполнением процесса практически в реальном времени, но и вносить изменения и почти мгновенно видеть результат. Значительно сокращается время между обнаружением проблемы, разработкой решения и оценкой эффекта от этого решения. В результате процессы и аудит выполняются синхронно, а решения принимаются здесь и сейчас на основании самых актуальных данных.

Однако, помимо плюсов, внедрение ИИ само по себе создает дополнительные риски, связанные с заложенными в модели вредоносными алгоритмами или потенциальными утечками конфиденциальных данных, что следует также учитывать при внедрении и последующей разработке соответствующих подходов и процедур.

Задавшись целью трансформации внутреннего аудита в условиях развития ИИ, компании переходят от практики проведения проверок к разработке интеллектуальных систем контроля и мониторинга. На смену статичному ретроспективному анализу, основанному на выборке, приходит динамичный мониторинг, моделирование потенциальных рисков и опасных ситуаций, реактивная модель сменяется предиктивной.

Расширение компетенций аудитора

Залог успешного перехода к ИИ-аудиту и внедрения передовых цифровых инструментов – квалифицированные сотрудники с широким набором кросс-функциональных компетенций. Новые подходы требуют от аудитора не только владения профессиональными стандартами, но и навыками, связанными с обработкой и анализом данных.

Аудитор становится гибридным специалистом: это и практик, который понимает бизнес и его риски, и аналитик, уверенно использующий SQL, Python и BI-инструменты, и промпт-инженер, качественно формулирующий запросы, позволяющие извлечь максимум значимых и полезных сведений из данных и документации. Он способен выявлять галлюцинации в ответах ИИ, поскольку обладает широким спектром практических знаний в разных областях бизнеса.

Кроме того, аудит должен способствовать развитию кросс-функционального обучения в компании. Аудиторы учатся говорить на одном языке с ИТ-специалистами и транслируют

эти знания другим сотрудникам. Аудиторы учатся оценивать качество данных и улучшать его, минимизируя риск «мусор на входе – мусор на выходе» (GIGO, garbage in, garbage out). Аудиторы работают с ИИ, но также проходят специализированное обучение для того, чтобы лучше понимать его возможности и изъяны, алгоритмы и механизмы функционирования.

Довольно часто аудитору приходится изучать и анализировать достаточно большой объем документации. Специально разработанный и обученный ИИ-агент может взять на себя эту функцию и кратно увеличить скорость и объем анализа. Грамотно обученный цифровой помощник способен провести многофакторный анализ и предложить возможные гипотезы и решения.

Безопасная среда для тестирования и экспериментов

Одним из главных барьеров внедрения ИИ в аудите является риск утечки данных или нарушения работы критических систем. Оптимальным решением будет создание изолированной «песочницы» для обучения и тестирования новых инструментов.

«Песочница» дает аудиторам возможность:
- осваивать новые инструменты без рисков для деятельности компании;
- тестировать новые алгоритмы и выбирать наиболее оптимальные;
- проверять гипотезы и настраивать модели на образцах реальных данных, не затрагивая при этом действующие информационные системы компании;
- имитировать угрозы и проверять возможности и устойчивость защитных систем;
- разрабатывать и тестировать ИИ-инструменты для решения отдельных задач и автоматизации аудиторских и аналитических процедур.

«Песочница» становится лабораторией экспериментов и разработок, средой для тестирования и обучения. Здесь можно сравнить различные решения и инструменты, выбрать наиболее эффективные и безопасные и предложить их бизнес-подразделениям для внедрения и использования. Так внутренний аудит может стать технологическим лидером и проводником изменений в компании, при этом не затрагивая действующую контрольную среду.

Внедрение ИИ: от пилотирования к системе

Как и в случаях с другими информационными системами, внедрение ИИ в аудит требует системного подхода. Качественная подготовка позволяет минимизировать или предотвращать возникновение большинства рисков. Перед этим стоит оценить внутренний ландшафт компании:
- аудит данных: оценить полноту достоверность, доступность данных в основных процессах;
- расширение компетенций аудита: включение в команду аудиторов специалистов по работе с данными и ИИ-моделями;
- обучение: организация специализированного обучения аудиторов по направлениям работы с данными и ИИ-моделями;
- организация «песочницы» и поддержка внутренней разработки ИИ-агентов: создание безопасной среды для обучения и экспериментов;
- пилотирование: проверка эффективности и работоспособности инструмента на одном процессе;
- масштабирование: внедрение эффективных практик в подразделениях компании;
- мониторинг: регулярная проверка алгоритмов на корректность работы.

Важным фактором успешного внедрения является поддержка нововведений сотрудниками. Помимо технических решений, значительное влияние на успех оказывают вовлеченность и заинтересованность людей, которые с этими решениями работают. Поскольку внутренний аудит становится технологическим проводником, от эмоционального настроя и энтузиазма аудитора

будет зависеть успешность трансформации процессов и применения новых технологий. Аудитор из контролера превращается в партнера, наставника.

Адаптация существующих на рынке ИИ-решений в совокупности с самостоятельной разработкой ИИ-инструментов позволит создать эффективную и безопасную экосистему, учитывающую особенности и специальные потребности компании.

Преимущества использования ИИ-инструментов в аудите

ИИ-инструменты позволяют внутреннему аудиту выйти на новый уровень мониторинга и контроля процессов в компании. В первую очередь – за счет возможности быстро и эффективно анализировать 100% объема данных. Это позволяет выявлять все проблемные области и нарушения, а впоследствии – предотвращать их появление. Мониторинг данных может проводиться в режиме реального времени, что, в свою очередь, позволит максимально оперативно реагировать на выявленные угрозы. Автоматизация части рутинных процедур избавит аудитора от необходимости долго и тщательно изучать документацию. Таким образом, без увеличения нагрузки на сотрудника повысится качество и увеличатся охват и глубина проверок.

Аудит ИИ-инструментов

ИИ-инструменты и инструменты, разработанные на базе или с использованием ИИ, также должны пройти надежную проверку на безопасность. Аудит кода, оценка соответствия решения техническому заданию, отсутствие «закладок» и шпионских модулей, корректность моделей – все это можно проверить с помощью специально обученных ИИ-моделей. Более того, с помощью одной ИИ-модели можно проверить другую ИИ-модель. Раньше такая проверка потребовала бы привлечения множества узких специалистов и заняла бы очень много времени, а сейчас новые технологии позволяют сократить затраты и повысить точность проверки.

Выводы

Когда мы говорим про ИИ, очень важно осознавать, что мы говорим про сегодняшний день. Это реальность, без адаптации к которой устоявшиеся подходы к работе быстро теряют актуальность и ценность.

Внутренний аудит не может и не должен оставаться в стороне. В ответ на новые вызовы меняется его роль: он становится проводником инноваций и новых методов, когда в основе прогнозов и рекомендаций лежат реальные данные, а технологии способствуют прозрачности, управляемости и развитию компании.

Но также не стоит забывать, что центром системы остается человек. Именно он обучает модели, выбирает и проверяет ИИ-инструменты, определяет наиболее оптимальный сценарий применения технологий и в конечном счете принимает решение на основании полученных данных. Он объединяет все инструменты в единую экосистему, задает правила и контролирует их соблюдение. Опираясь на свои знания и компетенции, он помогает компании эффективно справляться с вызовами завтрашнего дня.

Список источников

1. Исследование текущего состояния и тенденций развития внутреннего аудита в России // Институт внутренних аудиторов. М. 2025 (дата обращения: 15.07.2026).
2. Филатова Е.В., Выжевская И.А. Применение технологий искусственного интеллекта в практике государственного аудита // Парадигма. 2025. № 10.1 (дата обращения: 15.07.2026).
3. Шевченко А.А., Юрасова И.О. Использование информационных систем и искусственного интеллекта во внутреннем аудите и контроле // Экономика и бизнес: теория и практика. 2025. № 3 (121) (дата обращения: 15.07.2026).
4. Ahmi A., Smidt L. AI-driven auditing: trends, themes, and research trajectories. 2026 (дата обращения: 15.07.2026).
5. Barbu A.-G. The hidden risks of generative artificial intelligence in accounting and auditing: a bibliometric perspective // Annals of Faculty of Economics. 2025. Vol. 34, no. 2. P. 336–346 (дата обращения: 15.07.2026).

Ознакомиться с презентацией

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ОПЕРАЦИОННЫХ ПРОЦЕССАХ X5: ОСНОВНЫЕ РИСКИ

Финансовые метрики для расчета окупаемости ИИ-инструментов не отличаются от других инвестиционных проектов, требующих капитальных вложений, равно как и пороговые значения этих метрик. Из-за дорогостоящего оборудования замещение офисных сотрудников ИИ-агентами не даст необходимого уровня окупаемости, поэтому гораздо важнее их использование в областях, которые помогут компании больше зарабатывать.

И.Н. ВОРОНА
Директор по внутреннему аудиту ПАО «Корпоративный центр ИКС 5»

Аннотация

В статье описывается практический опыт применения ИИ в операционной деятельности крупной розничной компании, включая использование ИИ-агентов и машинного обучения для автоматизации рутинных задач, подходы к контролю и внутреннему аудиту ИИ-решений. Также внимание уделяется оценке экономического эффекта от внедрения ИИ, митигации рисков, связанных с внедрением ИИ.

Ключевые слова: искусственный интеллект, аудит ИИ-решений, цифровая трансформация, автоматизация в ретейле, ИИ-агенты.

Коды JEL: O33, D83, L81, L86, M11.

Внедрение ИИ в группе X5 – это часть стратегии до 2028 года, согласно которой компания намеревается достичь существенного прогресса при трансформации в самого технологичного омниканального ретейлера в мире. Помимо роботизации перемещения товаров и автоматизации администрирования/управления, это включает в себя ИИ-аугментацию офисных ролей. ИИ-аугментация означает использование офисными сотрудниками компании ИИ-инструментов в качестве помощников, которые усиливают, расширяют и дополняют человеческие возможности. А к ИИ-инструментам в компании относят не только ИИ-агентов с использованием больших языковых моделей (LLM), но и технологии машинного обучения (ML).

В самом подходе к внедрению ИИ в компании уже заложен механизм митигации рисков, связанных с ошибочными действиями ИИ-агентов и галлюцинацией больших языковых моделей. Поскольку LLM-модели дают хороший результат при обработке массовых повторяющихся рутинных операций (алгоритмы обработки которых улучшаются с помощью ML), именно такие задачи передаются ИИ-агентам, чтобы разгрузить сотрудников. Однако критическая оценка предлагаемых ИИ-решений и последнее слово остаются за сотрудником.

Во внедрении ИИ в компании, можно выделить три этапа: подготовка инфраструктуры и сотрудников, пилотирование ИИ-агентов силами бизнес-подразделений с замером экономического эффекта, масштабирование и встраивание ИИ-агентов в бизнес-процессы.

На подготовительном этапе в 2025 году был запущен продукт CoPilot X5, который через веб-интерфейс дает доступ сотрудникам к передовым открытым ИИ-моделям (его сейчас ежемесячно используют более 5 тыс. сотрудников), таким как ChatGPT, GigaChat, YandexGPT, а также Ai-Run X5 (установлена в закрытом контуре корпоративной ИТ-сети) для работы с конфиденциальными данными. Это помогло

снять психологический барьер у сотрудников, возникающий при обращении к ИИ впервые, а также позволил решить проблему теневого использования ИИ, когда сотрудники в тайне от работодателя могут обрабатывать информацию, содержащую коммерческую тайну, в открытых внешних моделях.

В 2025 году был внедрен собственный новый высокопроизводительный GPU-кластер (Graphics Processing Units) как единая вычислительная ферма для работы с ИИ. Также был разработан конструктор ИИ-агентов, с помощью которого сотрудники бизнес-подразделений, прошедшие специальное обучение, могут сами создавать ИИ-агентов для своей работы. Запуск платформы для A/B тестирования эффектов и оценки их статистической значимости позволил замерять финансовый эффект от внедрения ИИ-агентов (необходимый для расчета окупаемости инвестиций в ИИ-инструменты).

В 2026 году в компании активно внедряются цифровые платформы для реинжиниринга, унификации и автоматизации процессов: по управлению персоналом, а также коммерческих, логистических и других процессов. В процессы встраиваются ИИ-агенты для выполнения массовых рутинных задач.

Торговые сети, бизнес-единицы и корпоративные функции X5 в настоящее время пилотируют использование ИИ-агентов на различных направлениях деятельности компании с одновременным замером экономического эффекта. В качестве примеров успешного применения ИИ в операционных процессах компании можно привести следующее: создание персональных предложений для покупателей торговой сети «Пятерочка» на основе данных о покупках в прошлом; проверка договоров аренды на соответствие шаблонам, при которой ИИ находит нестандартные формулировки и подсвечивает юристам потенциальные риски для окончательного утверждения правок; контроль на фабриках-кухнях с помощью компьютерного зрения, моет ли персонал руки должным образом.

Стоит упомянуть вопрос окупаемости ИИ-агентов. Финансовые метрики для расчета окупаемости ИИ-инструментов не отличаются от других инвестиционных проектов, требующих капитальных вложений, равно как и пороговые значения этих метрик. Из-за дорогостоящего оборудования замещение офисных сотрудников ИИ-агентами не даст необходимого уровня окупаемости, поэтому гораздо важнее их использование в областях, которые помогут компании больше зарабатывать, например за счет автоматического определения полочной цены товаров или благодаря прогнозированию спроса на товары для целей пополнения запасов.

В следующем году в X5 предполагается масштабирование использования ИИ-агентов, исходя из соотношения затрат и экономического эффекта от внедрения. В этот период будет также целесообразно провести внутренний аудит непосредственно процесса внедрения ИИ-инструментов, включая оценку степени митигации ИИ-рисков, а также начать проверять работу отдельных ИИ-агентов, встроенных в бизнес-процессы.

Поскольку любая самая совершенная ИИ-модель может стать источником риска, важно своевременно выявлять и устранять подобные риски. К основным группам рисков, связанных с использованием ИИ, необходимо отнести:

- риск утечки конфиденциальных данных, обрабатываемых с помощью ИИ;
- риски, связанные с вычислительными мощностями, необходимыми ИИ: (1) доступность/нагрузка (влияет на время отклика и, соответственно, производительность сотрудников, обращающихся к ИИ-агентам); (2) окупаемость инфраструктуры при использовании ИИ взамен труда сотрудников;
- риск ошибочного действия/решения ИИ-агента, в том числе из-за галлюцинаций.

Остановимся на последнем риске. Не стоит ждать, когда ИИ избавится от галлюцинаций. Их надо воспринимать как неотъемлемую черту ИИ, когда он ищет не истину, а создает правдоподобный результат. На практике нельзя на 100% отследить, что и как делает ИИ-агент, но без должного контроля он может превратиться в черный ящик. Сделать работу ИИ более прозрачной можно за счет последовательного применения нескольких ИИ-агентов, каждый из которых отрабатывает отдельную задачу. В таком случае сотрудникам легче критически оценить результаты работы каждого ИИ-агента и исправить их до того, как следующий агент будет отрабатывать дальнейшую задачу. Например, во внутреннем аудите при оценке дизайна внутреннего контроля можно использовать несколько специальных ИИ-агентов для описания бизнес-процесса, выявления в процессе существующих и отсутствующих элементов контрольных процедур, заполнения матрицы рисков и контроля, формулировки недостатков во внутреннем контроле. При последовательном применении таких ИИ-агентов внутренний аудитор может своевременно осмыслить и исправить созданные ИИ аудиторские документы, так как ответственность за их качество лежит на специалисте.

Внедрение ИИ в группе X5 снимает рутинную нагрузку с сотрудников и повышает эффективность операционных процессов при одновременном сохранении за человеком окончательного решения. Практическая реализация включает подготовку инфраструктуры и персонала (CoPilot X5, закрытый Ai-Run X5, собственный GPU-кластер), пилотирование с измерением экономического эффекта и последующее масштабирование только там, где капиталовложения и эксплуатационные расходы обеспечивают приемлемую окупаемость. В части митигации ИИ-рисков компания сосредоточена на предотвращении утечек конфиденциальных данных, обеспечении доступности вычислительных мощностей и снижении последствий ошибочных решений ИИ-агентов.

Список источников

1. Годовой отчет X5 за 2025 год (дата обращения: 15.07.2026).
2. Неверов М. X5 встал на путь ИИ-агентизации // TAdviser. 2026 (дата обращения: 15.07.2026).

Ознакомиться с презентацией

ЗАКЛЮЧЕНИЕ

С расширением сферы применения ИИ возрастают требования к качеству управления модельными рисками и ИИ-рисками, внутреннему контролю и внутреннему аудиту.

Материалы настоящего выпуска демонстрируют, что современный аудит моделей выходит далеко за рамки проверки отдельных алгоритмов. Все большее значение приобретает оценка системы управления модельным риском в целом: процессов разработки, качества данных, валидации, мониторинга, распределения ответственности и контроля изменений на протяжении всего жизненного цикла модели.

Опыт, представленный авторами, показывает, что эффективное применение ИИ-технологий невозможно без тесного взаимодействия специалистов по внутреннему аудиту, управлению рисками, ИБ, ИТ и бизнеса. Именно такой междисциплинарный подход позволяет обеспечить надежность моделей и повысить качество принимаемых управленческих решений.

Новый формат подготовки заседаний, основанный на проведении тематических мастер-классов, подтвердил свою востребованность. Он позволил существенно расширить профессиональную дискуссию, объединить опыт представителей различных организаций и повысить практическую ценность опубликованных материалов. Этот подход планируется использовать и при подготовке следующих выпусков дайджеста.

Редакционная коллегия надеется, что представленные материалы будут полезны специалистам в области внутреннего аудита, внутреннего контроля и управления рисками, руководителям организаций, членам советов директоров и комитетов по аудиту, а также всем, кто заинтересован в развитии современных подходов к безопасному и эффективному применению ИИ-технологий.

ЭКОСИСТЕМА ЭКСПЕРТНОГО СОВЕТА

При Банке России действует Экспертный совет по регулированию, методологии внутреннего аудита, внутреннего контроля и управления рисками. Он работает на постоянной основе и объединяет представителей Банка России, органов государственной власти, финансовых организаций, компаний реального сектора, научного сообщества и профессиональных объединений.

Задачи Экспертного совета – развивать профессиональное сообщество, готовить методические и аналитические материалы, а также вырабатывать подходы к совершенствованию внутреннего аудита, внутреннего контроля, управления рисками, в том числе модельными рисками и рисками, связанными с применением ИИ.

Ключевые элементы экосистемы

Экспертный совет – это, по сути, площадка, где встречаются регулятор, рынок, бизнес и наука, обсуждаются актуальные вопросы регулирования и формируются предложения по развитию профессии. Здесь рождаются лучшие практики внутреннего аудита, внутреннего контроля и управления рисками, а затем их подхватывает уже все профессиональное сообщество.

Внутри Экспертного совета постоянно функционируют рабочие группы. Именно они разрабатывают конкретные подходы и методологию по отдельным направлениям.

Регулярно проходят мастер-классы: ведущие эксперты разбирают реальные кейсы и делятся практикой применения новых технологий.

Заседания собирают всех участников экосистемы для обсуждения текущей повестки, а по их итогам готовятся методические и аналитические материалы, включающие рекомендации, обзоры, исследования.

О том, что происходит в Экспертном совете, можно узнавать из дайджестов, в которых собраны итоги заседаний, мастер-классов и главные профессиональные новости.

Профессиональное сообщество

Сегодня к сообществу присоединились уже более 500 экспертов из Банка России, госорганов, финансового и реального секторов, науки, профессиональных объединений, аудиторских и консалтинговых компаний.

Основой сообщества являются руководители и специалисты по внутреннему аудиту, внутреннему контролю, управлению рисками, комплаенсу и цифровой трансформации. Они обмениваются практиками, участвуют в рабочих группах и мастер-классах, предлагают идеи по развитию методологии.

Официальные ресурсы

Информацию о деятельности Экспертного совета можно найти на его официальной странице на сайте Банка России. Там публикуются положение Экспертного совета, анонсы и итоги заседаний, а также методические материалы – этот ресурс открыт для всех.

Для сотрудников Банка России есть еще один канал: внутренний информационный портал – сообщество в интранете. Здесь собраны материалы заседаний и мастер-классов, документы рабочих групп и архив публикаций.

Как присоединиться к сообществу

Проще всего через MAX или Telegram – по ссылкам или с помощью QR-кодов:

MAX

Telegram

Контакты

Если вы хотите принять участие в работе Экспертного совета, рабочих групп или предложить тему для мастер-класса, напишите нам: expert.board@mail.cbr.ru.

ГЛОССАРИЙ

Аудит моделей – направление внутреннего аудита, обеспечивающее независимую оценку корректности процессов разработки, валидации, внедрения и мониторинга моделей, а также управляемости их жизненного цикла в целом.

Аудит-трейл – журнал, фиксирующий действия и решения ИИ-системы (входные и выходные данные, примененные политики и ограничения, метаданные) и обеспечивающий возможность реконструкции и проверки принятых решений.

Внутренний аудит – независимая и объективная деятельность по предоставлению гарантий и консультаций, направленная на совершенствование деятельности организации и повышение эффективности процессов управления рисками, внутреннего контроля и корпоративного управления.

Внутренний контроль – процесс, осуществляемый руководством и сотрудниками организации, направленный на обеспечение надежности финансовой отчетности, эффективности операций и соблюдения требований законодательства и внутренних регламентов.

Галлюцинация модели – генерация ИИ-моделью фактически неверной, вымышленной или не подтвержденной источниками информации, которая выглядит правдоподобно.

Данные и управление данными – процессы обеспечения качества, целостности, доступности и надежности данных, используемых для принятия управленческих решений.

Жизненный цикл модели – последовательность этапов существования модели – от инициирования, сбора данных, разработки и валидации до внедрения, мониторинга, внесения изменений и вывода из эксплуатации (списания).

ИИ-агент – программная система, которая самостоятельно принимает решения о способе выполнения поставленной задачи и совершает действия (обращение к системам, отправка данных, взаимодействие с другими сервисами) без непрерывного участия человека.

ИИ-аугментация – подход к применению ИИ, при котором ИИ выступает в роли помощника, дополняющего и усиливающего возможности сотрудника, а не заменяющего его на рабочем месте.

ИИ-риск – подвид нефинансового риска, возникающий на стыке ИБ-рисков, ИТ-рисков, модельных, классических операционных, этических и регуляторных рисков, связанный с разработкой и применением ИИ-систем.

Искусственный интеллект (ИИ) – комплекс технологических решений, позволяющий имитировать когнитивные функции человека и получать результаты, сопоставимые с результатами интеллектуальной деятельности человека.

Карта рисков – инструмент визуализации и систематизации рисков организации, отражающий вероятность возникновения рисков и масштаб их потенциального воздействия.

Киберриски – риски, связанные с угрозами информационной безопасности, включая кибератаки, утечки данных и нарушение функционирования информационных систем.

Корпоративное управление – система взаимоотношений между советом директоров, исполнительным руководством и заинтересованными сторонами, направленная на эффективное управление организацией и защиту интересов собственников.

Модельный риск – риск ошибок процессов разработки, проверки, адаптации, приемки и применения методик количественных и качественных моделей оценки активов, рисков и иных показателей, используемых при принятии управленческих решений; рассматривается как вид операционного риска.

Объяснимость (explainability) – свойство модели или ИИ-системы, характеризующее возможность понятно для человека раскрыть логику формирования результата и обосновать принятое моделью решение.

Продуктовый риск-менеджмент – подход к управлению ИИ-рисками, при котором меры по минимизации рисков применяются заблаговременно: на этапах планирования, обеспечения данными, разработки и развертывания системы, до ее вывода в промышленную эксплуатацию.

Промпт-инъекция – вмешательство пользователя или третьего лица в запросы (промпты) к ИИ-модели с целью изменить ожидаемые результаты или функции модели в обход установленных ограничений.

Реестр моделей – документ (информационная система), содержащий полную информацию об используемых в организации моделях, включая их назначение, уровень риска и влияние на финансовый результат и организационную структуру.

Результат интеллектуальной деятельности (РИД) – охраняемый результат творческой деятельности (произведение, изобретение и другое), права на который регулируются частью четвертой Гражданского кодекса Российской Федерации.

Риск-аппетит – уровень риска, который организация готова принять для достижения своих стратегических целей.

Система внутреннего контроля (СВК) – совокупность организационных мер, процедур и механизмов, направленных на обеспечение надежности и эффективности деятельности организации.

Система управления рисками (СУР) – совокупность процессов, методов и инструментов, обеспечивающих выявление, оценку, мониторинг и управление рисками организации.

Сиротское произведение – результат, сгенерированный с использованием ИИ на основе объектов, правообладателя которых установить затруднительно, что создает риски нарушения интеллектуальных прав при его дальнейшем использовании.

Технологические риски – риски, связанные с использованием информационных технологий, цифровых платформ и технологических решений в деятельности организации.

Управление модельным риском (Model Risk Management, MRM) – совокупность процессов, обеспечивающих полный жизненный цикл модели, независимое распределение функций при ее создании, валидации и мониторинге, а также достаточность и документированность набора индикаторов и тестов, характеризующих качество модели.

Управление модельными рисками – процессы выявления, оценки и контроля рисков, возникающих при использовании математических моделей, алгоритмов машинного обучения и ИИ-систем.

Цифровая трансформация – процесс внедрения цифровых технологий, изменяющий бизнес-процессы, модели управления и организационные структуры компаний.

Цифровая устойчивость – способность организации поддерживать непрерывность деятельности и устойчивость бизнес-процессов в условиях технологических изменений и киберугроз.

Цифровое доверие (Digital Trust) – ожидание заинтересованных сторон, что цифровые технологии и организации, предоставляющие их, будут защищать интересы всех участников и соответствовать этическим и регуляторным ожиданиям общества; включает такие компоненты, как прозрачность, безопасность, конфиденциальность, справедливость и подотчетность.

Цифровой аудит – применение цифровых технологий, анализа данных и автоматизированных инструментов для проведения аудиторских проверок.

СПИСОК СОКРАЩЕНИЙ

Общие аббревиатуры и термины

ИБ – информационная безопасность

ИИ – искусственный интеллект

BI (Business Intelligence) – бизнес-аналитика

CPU (Central Processing Unit) – центральное процессорное устройство.

d-people (data professionals) – специалисты, обладающие компетенциями в области анализа данных и цифровых технологий

GenAI (Generative Artificial Intelligence) – генеративный искусственный интеллект

GPU (Graphics Processing Unit) – графический процессор

LLM (Large Language Model) – большая языковая модель

ML (Machine Learning) – машинное обучение

RAG (Retrieval-Augmented Generation) – генерация ответа с дополнением за счет поиска по внешним источникам данных

SQL (Structured Query Language) – язык структурированных запросов

Сокращения в области корпоративного управления, аудита и контроля

ПВР – подход на основе внутренних рейтингов (применяется при расчете величины кредитного риска)

СВА – служба внутреннего аудита

СУР – система управления рисками

MRM (Model Risk Management) – управление модельным риском

ModelOps – совокупность практик управления полным жизненным циклом модели

MLOps/MLSecOps – практики и инструментарий безопасной разработки, развертывания и эксплуатации моделей машинного обучения

RACI (Responsible, Accountable, Consulted, Informed) – модель распределения ответственности и полномочий участников процесса

Сокращения в области права

ГК РФ – Гражданский кодекс Российской Федерации

ТК РФ – Трудовой кодекс Российской Федерации

РИД – результат интеллектуальной деятельности

Международные организации и методологии

IIA (Institute of Internal Auditors) – международная профессиональная ассоциация внутренних аудиторов

COSO (Committee of Sponsoring Organizations of the Treadway Commission) – организация, разработавшая международные концепции внутреннего контроля и управления рисками

COSO ERM (Enterprise Risk Management Framework) – концептуальная модель управления рисками организации

ISO (International Organization for Standardization) – Международная организация по стандартизации

NIST (National Institute of Standards and Technology) – Национальный институт стандартов и технологий США

NIST CSF (Cybersecurity Framework) – рамочная модель управления кибербезопасностью

Экономическая классификация

JEL (Journal of Economic Literature Classification System) – международная система классификации научных исследований в области экономики