

ЗА ПРЕДЕЛАМИ КОНТРОЛЯ:

ТЕОРИЯ ПРЕДЕЛОВ УПРАВЛЕНИЯ ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ

Аналитический синтез

Оубек Khodjaev

Май 2026 · okhodjaev.com

Часть серии: «*Beyond Control: Theory of Limits of AI Governance*» (Эссе 1–12)

DOI: doi.org/10.5281/zenodo.20120514

АННОТАЦИЯ

Системы управления искусственным интеллектом (ИИ)^[1] не смогли обеспечить эффективный надзор не из-за некомпетентности регуляторов или недобросовестности разработчиков, а вследствие структурной неспособности самих институциональных архитектур, привлечённых для выполнения этой функции.

Настоящий синтез опирается на двенадцать аналитических эссе и формулирует единый структурный вывод: три ограничения — суверенный приоритет, инфраструктурная предопределённость и структурная несовместимость — одновременно воздействуют на любую архитектуру управления ИИ и усиливают друг друга: их взаимодействие носит мультипликативный характер. Суть не в том, что ограничения «усиливают» друг друга, а в том, что при совместном действии они делают правоприменение невозможным. Когда все три ограничения проявляются одновременно, система управления теряет способность не только предотвращать развитие проблемы, но и содержательно корректировать её последствия.

Из этого вывода следует не политическая дорожная карта, а смена самой точки отсчёта: от вопроса «как выстроить лучшее управление» к вопросу «какие механизмы ограничения ещё сохраняют значение после закрытия окна коррекции».

I. Введение: Рамка

В месяцы, предшествовавшие официальному роспуску Советского Союза в августе 1991 года, его управленческая машина оставалась операционно нетронутой. Комитеты заседали. Отчёты составлялись. Сессии проводились. Видимость контроля продолжалась бесперебойно — тогда как архитектура обратимости уже была демонтирована. Вопрос, не покидавший меня с тех пор, состоит не в том, почему система рухнула, — а в том, когда коррекция стала структурно невозможной, пока её видимость продолжалась без какого-либо признания этого факта.

Именно этот вопрос составляет рамку данного документа.

Доминирующий дискурс в области управления ИИ^[1] трактует проблему как вопрос проектирования: какой институт должен располагать полномочиями, какой механизм прозрачности должен быть предписан, какой технический стандарт должен быть установлен.

Данный документ отстаивает иную рамку.

Проблема не в том, что никто ещё не разработал правильную архитектуру управления. Проблема в том, что три структурных ограничения определяют пространство, в рамках которого любая архитектура должна функционировать, — и в этом пространстве условия реального правоприменения^[2] не могут быть выполнены одновременно.

Это не пессимистическое утверждение. Это структурное.

Теорема о пределах — не предписание к бездействию, это репозиционирование проблемы. Те, кто продолжает продвигать архитектуры управления, не признавая три структурных ограничения, не просто заблуждаются — они расходуют остаточный корректирующий потенциал, который ещё существует, поддерживая нарратив, делающий коррекцию кажущейся излишней — вплоть до момента, когда она становится невозможной.

Методологическое примечание.

Настоящий документ представляет собой аналитический синтез, основанный на тридцатипятилетнем опыте работы в банковском секторе, финансах, региональном управлении и бизнесе в Узбекистане и странах СНГ — включая опыт наблюдения институционального распада периода 1991 года и последующих системных трансформаций.

Автор является независимым исследователем и не аффилирован с институтами и организациями, чьи подходы, модели и архитектуры управления рассматриваются в данном анализе. Аналитическая база документа — практические наблюдения, институциональный опыт и многолетний мониторинг динамики развития передовых систем ИИ.

Практический опыт работы в условиях институциональных ограничений, кризисов управления и системных сбоев формирует особую аналитическую оптику. Она позволяет раньше распознать момент, когда институты продолжают воспроизводить процедуры, механизмы отчётности и формальные структуры контроля — уже после того, как способность системы к значимой коррекции фактически утрачена.

II. Диагностический блок: Шесть паттернов сбоя

Прежде чем перейти к теории пределов, необходимо установить диагностическую базу: шесть паттернов сбоя, регулярно воспроизводящихся в практике управления ИИ в реальном времени, — каждый с задокументированным историческим прецедентом. Это не шесть независимых

проблем. Это шесть наблюдаемых проявлений одной базовой динамики: институциональная архитектура, созданная для предшествующего технологического порядка, функционирует применительно к системе, для которой она не предназначалась.

Таблица 1: Шесть паттернов сбоя управления

ПАТТЕРН СБОЯ	ИСТОРИЧЕСКИЙ ПРЕЦЕДЕНТ	ПРОЯВЛЕНИЕ В СФЕРЕ ИИ (2026)	НАБЛЮДАЕМЫЙ СИГНАЛ
Перформативный контроль ^[3]	Советский Госплан: отчётность документировала целевые показатели, скрывая системные сбои	Обязательства Anthropic по безопасности против суверенного приоритета Пентагона; документация, не ограничивающая развёртывание	Разрыв между публичными гарантиями безопасности и отсутствием независимой верификации
Ловушка прозрачности	Рейтинги AAA до 2008 г.: избыток данных, породивший непрозрачность подотчётности	Карточки моделей и отчёты по стресс-тестированию моделей (моделирование атак и сценариев отказа) без независимого доступа к внутренним механизмам	Объём документации по безопасности растёт, возможности независимого аудита остаются нулевыми
Дилемма регулятора	Рынок капитала Узбекистана, 1994: регулятор выбирает легитимность в ущерб компетенциям и скорости	Закон ЕС об ИИ разрабатывается в цикле, измеряемом годами; развитие возможностей — месяцами	Частота пересмотров регулирования без изменения результатов правоприменения
Миф о согласовании	Добровольные обязательства COP: институциональные обязательства, подтачиваемые конкурентным давлением	0,0002% человечества задаёт ценности для глобально развёртываемых систем; расхождение задекларированных и операционных целей	Роспуск команд безопасности в ведущих лабораториях по мере сжатия сроков развёртывания
Колониальный паттерн	Кондиционирование МФ 1990-х: фреймворки кредиторов экспортированы как универсальные стандарты	Данные — из контекстов Глобального Юга, модели обучаются в Калифорнии, правила пишутся в Вашингтоне, Брюсселе и Лондоне	Отсутствие незападного представительства в разработке эталонных тестов безопасности и оценке моделей

ПАТТЕРН СБОЯ	ИСТОРИЧЕСКИЙ ПРЕЦЕДЕНТ	ПРОЯВЛЕНИЕ В СФЕРЕ ИИ (2026)	НАБЛЮДАЕМЫЙ СИГНАЛ
Замыкание паттерна	Кризис совета директоров OpenAI (2023): формальные полномочия и реальный центр решений разошлись; структура удержалась, контроль испарился	Занесение Anthropic в чёрный список Пентагоном (2026): граница безопасности разработчика оспорена суверенным актором	Разрыв между задекларированными возможностями управления и реально действующими механизмами коррекции

Эти паттерны не исчерпывают всей картины, однако отражают наиболее устойчиво воспроизводящиеся структурные конфигурации в спорах вокруг управления передовыми системами ИИ. Они не функционируют независимо друг от друга. Напротив, каждый следующий паттерн усиливает последствия предыдущего.

Перформативный контроль — то есть воспроизводство видимости контроля без способности реально влиять на систему — создаёт условия для ловушки прозрачности: как только видимость становится достаточной для формальной подотчётности, информационная асимметрия превращается в инструмент. Ловушка прозрачности, в свою очередь, включает дилемму регулятора: институт, неспособный прочесть реальный профиль риска подконтрольной системы, не может одновременно обеспечить понимание, скорость реакции и институциональную легитимность.

Дилемма регулятора воспроизводит миф о согласовании: когда ни одна внешняя сторона не способна верифицировать соответствие заявленных целей реальному операционному поведению системы, разрыв между ними остаётся незаметным вплоть до момента критического сбоя.

Миф о согласовании закрепляет колониальный паттерн: фреймворки, произведённые ограниченным кругом акторов, экспортируются как универсальные стандарты, тогда как общества, наиболее уязвимые к системным сбоям, фактически исключены из процесса их формирования.

Вместе эти паттерны образуют замкнутую структуру: момент, в котором разрыв между задекларированным и реальным контролем становится неотрицаемым, наступает не через добровольное раскрытие проблемы, а через сбой, который сама архитектура управления должна была предотвратить.

III. Проблема измерения: Окно коррекции

Если шесть паттернов описывают, что именно не работает, полезный анализ управления требует метрики. Концепция окна коррекции^[4] — интервала между появлением системы и

моментом, когда откат (возврат к предыдущему состоянию) становится операционно неосуществимым, — даёт такую метрику.

- Реальное правоприменение — в отличие от его видимости — сводится к трём условиям, ни одно из которых в настоящее время не существует в операционной форме в управлении ИИ.
- Полномочие на остановку^[5]: способность прекратить развёртывание — не после рыночного провала, а до него.
- Независимый доступ: прямое изучение поведения системы изнутри, без опоры на самостоятельно предоставленную документацию.
- Последствия за дезинформацию: архитектура ответственности, при которой разрыв между задекларированным и реальным профилем безопасности влечёт юридически и финансово значимые санкции.

Все три в настоящее время отсутствуют в управлении ИИ в операционной форме.

Полномочие на остановку в отдельных юрисдикциях существует формально, однако не реализовано как действующий механизм предварительного контроля. Ни один регулятор сегодня не обладает полномочиями обязательного предварительного одобрения передовых систем ИИ, сопоставимыми с процедурами допуска фармацевтических препаратов. Закон ЕС об ИИ предусматривает постмаркетинговый надзор, но не содержит обязательного механизма предварительной остановки, который разработчик не мог бы обойти через выбор другой юрисдикции.

Независимого доступа также не существует. Внутренние механизмы передовых моделей остаются недоступными для внешнего аудита. Ни один регуляторный орган в настоящее время не располагает технической возможностью оценивать возникающие свойства сложных моделей ИИ на уровне интерпретируемости, необходимом для содержательной проверки безопасности.

Последствия за дезинформацию требуют наличия измеримой базы, относительно которой ложное утверждение может быть доказано ложным. Однако согласованной метрики безопасности ИИ не существует. В результате граница между добросовестной неопределённостью и намеренной дезинформацией начинает размываться в условиях подлинной неопределённости самого знания о поведении системы.

Отсутствие всех трёх — не временный дефицит, ожидающий лучшего законодательства. Это структурная характеристика нынешней среды управления — подтверждённая в Международном докладе о безопасности ИИ 2026 года, видимая в паттерне роспуска команд безопасности в ведущих лабораториях по мере ускорения развёртывания и задокументированная в растущем разрыве между изощённостью документации по безопасности и механизмами независимой проверки её утверждений.

Окно коррекции не закрывается. В большинстве критических областей оно уже закрылось. Интеграция передовых моделей ИИ в облачные сервисные API — тиражируемые, глобально распределённые, программируемые миллионами нижестоящих приложений — создала сеть

зависимостей, прерывание которой ни один регулятор не может осуществить без каскадного сбоя тысяч организационных функций одновременно. Эта невозможность не законодательная по природе — она структурная.

IV. Механизм деградации: Трансфер агентности

Окно коррекции закрывается не сразу. Оно закрывается через процесс: прогрессивную миграцию полномочий по принятию решений от людей-операторов к автоматизированным системам. Эта миграция — трансфер агентности^[6] (то есть передача способности к самостоятельному принятию решений) — является механизмом деградации, делающим закрытие окна коррекции структурным, а не обратимым.

Трансфер агентности — не техническое событие. Это институциональный процесс. Организации не принимают единого осознанного решения передать контроль системе ИИ. Они накапливают небольшие операционные зависимости — в планировании, маршрутизации, оценке рисков, взаимодействии с клиентами, — каждая из которых оправдана повышением эффективности. Накопление пересекает порог: человеческая компетенция, необходимая для функционирования без системы, атрофируется вследствие неиспользования. В этой точке откат — не просто дорогостоящий. Он операционно неосуществим.

Механизм порождает то, что можно назвать деградацией реального контроля в момент исполнения^[7]: разрыв между задекларированным объёмом человеческого надзора и реальным принятием решений, происходящим в режиме автоматизации. Риск-офицер, способный остановить транзакцию, осуществляет реальный контроль. Риск-офицер, способный лишь задокументировать несогласие с рекомендацией системы, осуществляет перформативный контроль: документация фиксирует форму надзора, тогда как решение фактически принимает система. По мере углубления интеграции ИИ в критические функции организаций численность первой категории сокращается, второй — расширяется — не через политические решения, а через накопленный операционный факт. Ключевое отличие этого механизма от стандартных опасений об автоматизации: проблема не в том, что «ИИ забирает работу», — а в том, что человек деградирует и теряет саму способность управлять функцией без ИИ.

Эта динамика имеет особые последствия для институтов в юрисдикциях, получающих системы ИИ, которые они не проектировали. Их трансфер агентности происходит не только на организационном уровне, но и на суверенном: решения о возможностях, границах и циклах обновления критических систем принимаются акторами за пределами их территории, в рамках временных горизонтов, на которые они не могут влиять, под коммерческим и стратегическим давлением, которое они не контролируют. К примеру, унификация QR-платежей в Узбекистане (Постановление ЦБУ №3817, март 2026 года) — единая централизованная база поведенческих данных в режиме реального времени, контролируемая единым институциональным актором, без положения об откате в регулирующем документе — демонстрирует механизм в национальном масштабе: выгоды от эффективности реальны; риск концентрации институциональный, а не технический.

V. Теория пределов: Три структурных ограничения

То, что установили предшествующие разделы, — градиент сбоя без встроенного корректирующего механизма. Шесть паттернов институциональной деградации, сужающееся окно коррекции и прогрессивная атрофия агентности через накопленную зависимость — каждый ускоряется там, где сходятся три структурных ограничения. Вопрос, которому посвящён данный раздел, состоит в том, почему эти ограничения не могут быть преодолены даже в принципе — вне зависимости от качества институционального дизайна или политической воли.

Ответ состоит из трёх структурных ограничений. Каждое — структурный предел, а не устранимый дефицит. Их взаимодействие мультипликативно: одновременное действие всех трёх производит не среду, которую втрое труднее преодолеть, — а среду, в которой условия реального правоприменения не могут быть выполнены одновременно.

Ограничение первое: Суверенный приоритет

Каждая архитектура правоприменения в управлении ИИ встречает структурный предел в точке, где государство, для которого технология стала элементом стратегической автономии, отказывается подчинить свои интересы архитектуре управления. Суверенный приоритет^[8] (sovereign override) принадлежит структуре международного порядка, а не какому-либо просчёту в дизайне конкретного регуляторного фреймворка.

Спор Anthropic–Пентагон начала 2026 года иллюстрирует механизм в реальном времени. Частный актор (Anthropic) объявил границу безопасности. Суверенный актор (Пентагон), операционально зависящий от той же технологии, оспорил эту границу. Спор перешёл в федеральный суд. Второй суверен, наблюдая за конфликтом, открыл параллельный канал — британское правительство обратилось к Anthropic с предложениями о расширении деятельности вслед за американским конфликтом. Ни одна многосторонняя структура, профессиональный стандарт или отраслевая норма не функционировала как обязывающее ограничение. Единственными инструментами оставались договорное право, регуляторное определение, судебные разбирательства и конкурирующие интересы двух государств.

Перед нами воспроизводимый структурный паттерн, который проявляется во многих системах международного регулирования и контроля.

Договор о нераспространении ядерного оружия функционирует уже более пятидесяти лет при наличии государств, так и не присоединившихся к нему. Архитектуры финансовых санкций последовательно порождают параллельные платёжные системы по мере того, как государства формируют стимулы и возможности для обхода ограничений. Добровольные обязательства по безопасности ИИ воспроизводят ту же структурную логику, что и климатические соглашения по сокращению выбросов: они выглядят устойчивыми в момент подписания, но начинают размываться под давлением конкуренции, суверенных интересов и технологической гонки.

Сфера ИИ лишь радикально ускоряет этот цикл. Развёртывание технологий происходит в темпе релизных циклов, а не дипломатических процессов. Одновременно трансфер агентности постепенно атрофирует институциональные компетенции, необходимые для отката системы, — именно в тот момент, когда суверенные ставки становятся максимальными.

Ограничение второе: Инфраструктурная предопределённость

Второе ограничение действует через материю, а не через волю. Оно не преодолевается никаким суверенным инструментом, поскольку не подчиняется ни одному суверенному инструменту. Инфраструктурная предопределённость^[9] — физическая конфигурация вычислительного стека, не поддающаяся регуляторному контролю.

Оборудование для экстремальной ультрафиолетовой (EUV) литографии, производство передовых полупроводников, энергетическая инфраструктура для обучения моделей в фронтирном масштабе, география центров обработки данных — всё это предопределяет диапазон стратегических выборов, доступных большинству юрисдикций, до начала любого диалога об управлении ИИ. Это ограничение физическое, а не политическое.

Один завод в Нидерландах производит машины для EUV-литографии, без которых производство передовых чипов на ведущих технологических нормах невозможно. Одна компания на Тайване производит подавляющее большинство чипов фронтирного класса, на которых обучаются ведущие модели ИИ. Энергетические требования для обучения передовых моделей ИИ превышают мощность электросети большинства государств. Ни один регуляторный инструмент не достигает этих фактов. Они образуют материальный субстрат среды управления ИИ.

Для юрисдикций, не производящих технологию и не контролирующих вычислительную инфраструктуру, суверенитет в политике в области ИИ означает нечто структурно отличное от того, что суверенитет означает в других областях. Государство может издать постановление. Оно не может издать производственную линию. Декларация суверенитета и материальная возможность его реализации — это две разные вещи. В области ИИ разрыв между ними структурный — и инфраструктурная предопределённость ограничивает суверенный выбор до начала какого-либо обсуждения.

Ограничение третье: Структурная несовместимость

Третье ограничение — нехватка институциональных возможностей. Это структурная несовместимость^[10] между устройством существующих институтов управления и тремя элементами реального правоприменения.

Существующие институты были построены на допущении о легитимности публичного права в пределах определённой юрисдикции, технической доступности регулируемой системы и юрисдикционной согласованности между масштабом проблемы и масштабом полномочий.

Управление ИИ не требует, чтобы ни одно из этих условий выполнялось так, как институты спроектированы для их выполнения.

Разработчики ИИ находятся в одной юрисдикции, развёртывание происходит в десятках, а последствия затрагивают сотни других государств. Ни один регулятор не обладает полномочиями одновременно контролировать все три фазы этой цепочки: разработку, развёртывание и последствия.

ЕС способен регулировать использование ИИ внутри собственных границ, но не разработку используемых систем. США могут ограничивать экспорт отдельных вычислительных ресурсов, однако не способны контролировать развёртывание архитектур, уже ставших глобально распределёнными.

При этом ни один регуляторный орган сегодня не располагает технической возможностью оценивать возникающие свойства передовых моделей ИИ на уровне интерпретации их внутренних механизмов, необходимом для содержательной проверки безопасности. И речь идёт не о временном отставании. Разрыв продолжает ускоряться: цикл развития технологических возможностей измеряется месяцами, тогда как цикл институциональной адаптации — годами.

Мультипликативное взаимодействие: Формальная формулировка теоремы

Когда суверенный приоритет, инфраструктурная предопределённость и структурная несовместимость действуют одновременно на архитектуру управления ИИ, её корректирующая способность не снижается постепенно. Она обрушивается ниже порога, при котором значимая коррекция остаётся операционно осуществимой — не потому, что институты некомпетентны, а потому что структурные условия, необходимые для реального правоприменения, не могут быть выполнены одновременно при нынешних геополитических и институциональных условиях. Это теорема пределов, а не прогноз. Она описывает архитектуру проблемы, а не конкретный исход.

Инфраструктурная предопределённость ограничивает суверенный выбор до начала обсуждения: государство не может принять позицию в области управления ИИ независимо от вычислительной инфраструктуры, контролируемой иностранными акторами.

Суверенный приоритет активируется, когда государство определяет, что его стратегический интерес перевешивает издержки соответствия, — но порог этого определения снижается именно вследствие инфраструктурной зависимости, поскольку государство, не способное производить альтернативы, располагает меньшими возможностями для применения границ.

Структурная несовместимость означает, что регуляторные инструменты, доступные для решения обоих предыдущих ограничений, требуют одновременного выполнения условий — полномочия на остановку, независимого доступа, последствий за дезинформацию, — которые взаимодействие первых двух ограничений делает недоступными.

VI. Остаточный управленческий ресурс

Теорема пределов не означает исчезновение всех механизмов ограничения. Она означает конец архитектуры управления^[12], основанной на предположении, что три структурных ограничения могут быть преодолены за счёт более совершенного институционального дизайна. То, что остаётся, — остаточный управленческий ресурс^[11] (governance residual): совокупность механизмов ограничения, сохраняющих реальный операционный рычаг даже после того, как окно коррекции для системного управления передовыми системами ИИ уже закрылось.

Остаточный управленческий ресурс включает три категории. Ни одна из них не заменяет отсутствующие элементы реального правоприменения. Однако каждая сохраняет значение там, где три структурных ограничения ещё не перекрыли весь доступный рычаг воздействия.

Страхование и андеррайтинг.

Когда крупные страховые компании начнут оценивать глубину зависимости от ИИ, как отдельный фактор андеррайтинга — отдельно от общего киберриска, — они создадут финансовый стимул для организаций сокращать зависимость прежде, чем это потребуют обязательные регуляторные архитектуры управления. Когда это расхождение появится в практике андеррайтинга, тогда как публичные гарантии безопасности от разработчиков останутся неизменными, это станет наиболее прямым рыночным сигналом о том, что корректирующая способность деградировала за пределы того, что отражает формальное управление.

Судебная практика и судебный реестр.

Когда судебные разбирательства, связанные с ИИ, систематически обязывают к раскрытию внутренних оценок безопасности, они создают судебные реестры, которые могут противоречить добровольным материалам о прозрачности. Суды не могут управлять ИИ в смысле надзора до развёртывания. Однако они могут быть механизмом, через который раскрытие информации в ходе разбирательства создаёт подотчётность за разрыв между задекларированным и реальным профилем безопасности — постфактум, но с последствиями, которые формальное управление в настоящее время произвести не может.

Смежные регуляторные полномочия.

Требования к закупкам, условия финансового лицензирования, архитектуры защиты данных и трудовое законодательство — каждое охватывает смежные с ИИ области, где формальное управление ИИ не функционирует. Юрисдикция, обуславливающая государственные закупки демонстрируемой способностью к откату — не декларируемым соответствием требованиям надзора, а реально продемонстрированной человеческой способностью выполнять критические функции без системы, — осуществляет управление через канал, который три ограничения не перекрыли.

Остаточный управленческий ресурс — не замена архитектуре управления, которую три ограничения сделали недостижимой. Это честная отправная точка для тех, кто действует после

закрытия окна коррекции. Использование остатка на поддержание нарратива — на сохранение видимости контроля — истощает последний доступный корректирующий потенциал. Выбор между признанной зависимостью и непризнанным коллапсом — это вопрос управления, остающийся открытым.

Структурная позиция Глобального Юга

Юрисдикции, находящиеся в структурной позиции Глобального Юга^[14], сталкиваются с этой проблемой раньше и острее других.

- Они потребляют системы ИИ, не участвуя в их проектировании;
- Подчиняются архитектурам управления без представительства в их формировании;
- Несут последствия системных сбоев без контроля над производящими их технологиями;
- Обладают формальным суверенитетом без материальной возможности его полноценной реализации.

В таких условиях окно коррекции закрывается быстрее, поскольку отсутствуют институциональные буферы, позволяющие центральным акторам дольше поддерживать видимость контроля.

Это не только ограничение. В определённом смысле это и аналитическое преимущество. Министерство закупок, признающее невозможность независимого аудита внедряемой системы ИИ, по крайней мере имеет возможность инвестировать в ручные резервные процессы и диверсификацию поставщиков. Министерство, продолжающее исходить из предположения о существовании полноценного надзора, с высокой вероятностью не сделает ни того, ни другого — и столкнётся с операционным разрывом уже после отказа системы в условиях, не предусмотренных документацией.

Те, кто не может долго поддерживать иллюзию контроля, раньше сталкиваются с остаточным управленческим ресурсом как с единственно доступной формой ограничения. Такое столкновение — без иллюзий относительно полноты контроля — и становится первой реалистичной точкой отсчёта для управления ИИ в условиях структурных ограничений.

VII. Три императива репозиционирования

Теорема пределов не производит политических рекомендаций. Она производит репозиционирование проблемы. Три структурных императива вытекают из анализа для любых институтов, министерств, банков и других организаций, принимающих решения о развёртывании ИИ или управлении им.

Императив 1: Аудировать глубину трансфера агентности, а не соответствие формальным требованиям.

Диагностический тест здесь предельно конкретен: проводила ли ваша организация в течение последних шести – двенадцати месяцев учения по неавтоматизированному откату критических функций? Если нет, существующая документация фиксирует не контроль, а уязвимость.

Разница принципиальна:

- соответствие требованиям измеряет наличие формальных механизмов надзора;
- глубина трансфера агентности показывает, остаётся ли сам надзор операционно осуществимым в момент сбоя.

Организация, измеряющая только первое, почти неизбежно столкнётся с разрывом между формой и содержанием в наихудший возможный момент: когда система откажет в условиях, не предусмотренных документацией, а человеческие компетенции, необходимые для отката, уже атрофировались вследствие длительного неиспользования.

Соответствие требованиям, не подтверждённое способностью к реальному откату системы, — не управление. Это нарратив.

Императив 2: Инвентаризировать остаточный управленческий ресурс прежде, чем брать обязательства в рамках театра соответствия.

Остаточный управленческий ресурс — страховая уязвимость, судебные механизмы, вето на закупки, смежные регуляторные полномочия — всегда ограничен. Институты, расходующие его на поддержание нарратива контроля, теряют его без реального ограничивающего эффекта.

Иное происходит там, где этот ресурс направляется на практические рычаги воздействия:

- независимый аудит как условие страхового покрытия;
- демонстрация способности к откату как условие закупок;
- обязательное судебное раскрытие как условие урегулирования.

Даже после того, как формальная архитектура управления становится операционно нерелевантной, такие механизмы всё ещё способны сохранять корректирующее влияние.

Ключевой диагностический вопрос заключается не в наличии архитектуры управления как таковой. Он заключается в другом: какие элементы остаточного управленческого ресурса реально используются для ограничения рисков, а какие — для поддержания театра соответствия^[13]? Именно это различие определяет, сохраняет ли институт остаточную способность к коррекции или постепенно расходует её.

Императив 3: Рассматривать архитектуры управления как свидетельство обязательств, а не как шаблоны.

Архитектуры управления ИИ, формирующиеся в центре экосистемы — Закон ЕС об ИИ, американские исполнительные указы, добровольные обязательства по безопасности, —

создаются прежде всего как ответ на вопрос: «Как должно выглядеть управление ИИ?» Они оптимизированы для институциональной легитимности, а не для способности к содержательной коррекции.

Архитектура управления, построенная вокруг поддержания видимости контроля, будет продолжать воспроизводить отчётность о намерениях даже после закрытия окна коррекции.

Юрисдикции, воспринимающие такие архитектуры как универсальные шаблоны, принимают требования к демонстрации контроля без приобретения самих корректирующих возможностей — которых в исходной модели зачастую и не существовало.

Поэтому архитектуры управления, разработанные в других центрах силы, следует рассматривать прежде всего как свидетельство того, к каким обязательствам центр готов публично присоединиться. Это почти всегда отличается от того, какие механизмы он действительно готов исполнять в условиях стратегического давления.

Практическое проектирование должно строиться не от идеальной архитектуры к реальности, а от реально доступных ограничивающих рычагов — наружу. Иными словами: от остаточного управленческого ресурса к архитектуре управления, а не наоборот.

VIII. Заключение: От управления к позиционированию

Серия началась с вопроса, заданного в Ташкенте в 1991 году, — о том моменте, когда откат стал структурно невозможным, пока видимость контроля продолжалась. Она завершается структурным ответом.

Институциональный порядок, созданный для управления предыдущими технологиями, структурно неспособен управлять передовыми системами ИИ — не потому, что люди внутри него некомпетентны или нечестны, а потому что три ограничения взаимодействуют мультипликативно, создавая среду управления, в которой условия реального правоприменения не могут быть выполнены одновременно.

Данный фреймворк не утверждает, что значимые вмешательства в управление ИИ невозможны в ограниченных областях или конкретных секторах. Он утверждает, что ни одна существующая архитектура не может одновременно выполнить условия, необходимые для системного управления передовыми системами ИИ при нынешних геополитических и институциональных условиях, — и что это справедливо вне зависимости от качества намерений или изощёренности институционального дизайна.

Определённые события и механизмы могут частично расширять окно коррекции в отдельных областях. К ним относятся:

- обязательные режимы ответственности, при которых разрыв между задекларированным и реальным профилем безопасности становится финансово значимым;

- системы лицензирования вычислительных мощностей, вводящие предварительные ограничения на уровне инфраструктуры;
- развитие интерпретируемости, обеспечивающее независимый доступ к поведению моделей;
- формирование суверенных вычислительных блоков, снижающих инфраструктурную предопределённость для участвующих юрисдикций.

Ни один из этих механизмов не устраняет три структурных ограничения. Однако каждый из них способен временно расширить остаточный управленческий ресурс в отдельных контекстах, создавая интервалы, в которых значимая коррекция ещё остаётся практически возможной для конкретных секторов или прикладных областей.

В этом и состоит теорема пределов. Её следствия носят не предписывающий, а позиционный характер.

Институты, проектирующие механизмы управления ИИ, должны исходить не из идеальных архитектур, а из реально сохраняющихся рычагов ограничения. Для организаций, управляющих зависимыми от ИИ операциями, ключевой вопрос состоит не в наличии архитектуры управления как таковой, а в том, сохраняются ли ещё человеческие компетенции, необходимые для отката системы — и что произойдёт, если ответ окажется отрицательным.

Для юрисдикций, находящихся в структурной позиции Глобального Юга, необходим иной подход: рассматривать архитектуры управления, разработанные в центре экосистемы, не как универсальные шаблоны, а как свидетельство того, к каким обязательствам центр готов публично присоединиться. Это почти всегда отличается от того, какие механизмы он действительно готов исполнять в условиях стратегического давления.

Окно коррекции не закрылось одновременно и окончательно во всех областях. Однако на структурном уровне оно уже закрылось для любой архитектуры, претендующей на полноценное системное управление передовыми системами ИИ.

То, что остаётся, — остаточный управленческий ресурс: частичный, асимметричный, нескоординированный и недостаточный для системной коррекции. Но остаётся и выбор — не между контролем и хаосом, а между признанной зависимостью и непризнанным институциональным коллапсом.

Те, кто продолжает действовать, как будто окно коррекции остаётся открытым, не просто заблуждаются. Они расходуют остаточный потенциал, который ещё существует, поддерживая нарратив, делающий коррекцию кажущейся излишней — вплоть до момента, когда она становится невозможной. То, что мы строим сейчас, не будет управлением ИИ. Это будет управление нашими собственными отношениями с системами, операционная непрерывность которых больше не зависит от значимой человеческой обратимости.

Примечания к терминологии

В настоящем аналитическом синтезе используются термины и концептуальные конструкции, сформированные автором в рамках серии эссе «Beyond Control» — фреймворка, посвящённого структурным ограничениям управления передовыми системами ИИ. Для читателей, не знакомых с предыдущими публикациями серии, ниже приведены краткие пояснения к ключевым понятиям, используемым в логике данного анализа.

[1] Система управления ИИ (AI Governance) — устоявшийся международный термин, обозначающий совокупность механизмов, норм, институтов и процедур, призванных обеспечить ответственное развитие и применение искусственного интеллекта. В данном документе переводится как «система управления ИИ» или «архитектура управления ИИ» в зависимости от контекста.

[2] Правоприменение (Enforcement) — реальное, а не декларируемое применение регуляторных требований с конкретными механизмами исполнения, независимой проверки и принудительных санкций. Принципиально отличается от соответствия требованиям (Compliance) — формального соблюдения норм без гарантии реального эффекта. В русскоязычном правовом дискурсе ближайший эквивалент — «правоприменение» или «принудительное обеспечение исполнения».

[3] Перформативный контроль (Performative control) — первый из шести диагностических паттернов фреймворка. Обозначает ситуацию, когда внешние формы контроля сохраняются и документируются, но его реальное содержание уже утрачено: производится видимость надзора, а не сам надзор. Аналог в русской административной традиции — «имитация контроля», однако авторский термин точнее передаёт перформативную (ритуальную) природу явления.

[4] Окно коррекции (Correction window) — ключевая концепция, введённая в Эссе 7 серии. Временной интервал, в течение которого коррекция системы остаётся операционно осуществимой: возможно остановить развёртывание, провести независимую проверку и применить санкции за дезинформацию. После закрытия окна эти действия становятся невозможными не по политическим причинам, а по структурным.

[5] Полномочие на остановку (Halt authority) — способность регулятора прекратить развёртывание системы ИИ на любом этапе — до, во время или после его начала. По аналогии с разрешительным порядком допуска лекарств к обороту в фармацевтической сфере. В настоящее время отсутствует для передовых систем ИИ во всех юрисдикциях.

[6] Трансфер агентности (Agency transfer) — введён в Эссе 8 серии. Агентность (Agency) — способность к самостоятельному иницированию действий и принятию решений. Трансфер агентности описывает постепенную миграцию этой способности от людей к

автоматизированным системам через накопление операционных зависимостей. Не следует путать с автоматизацией: ключевое явление — не замена людей машинами, а утрата людьми способности выполнять функцию без машины.

[7] Деградация реального контроля в момент исполнения (*Validity decay at execution time*) — введён в Эссе 8. Прогрессивный разрыв между формально задекларированным объёмом человеческого надзора и тем, что реально происходит в момент принятия решений автоматизированной системой. Показывает, как формальные процедуры надзора превращаются в ритуал документирования, тогда как фактические решения принимает система.

[8] Суверенный приоритет (*Sovereign override*) — первое из трёх структурных ограничений фреймворка. Термин *override* (англ.) означает активное преодоление или нейтрализацию действующей нормы в пользу иного приоритета. В данном контексте — способность и готовность государства игнорировать обязательства в рамках архитектуры управления ИИ, когда стратегические интересы требуют иного.

[9] Инфраструктурная предопределённость (*Material predetermination*) — второе из трёх структурных ограничений. Физическая конфигурация вычислительной, энергетической и производственной инфраструктуры (оборудование для EUV-литографии, полупроводниковое производство, энергосети), которая ограничивает пространство суверенных выборов в управлении ИИ до начала любых политических обсуждений. Ограничение действует через материю, а не через волю — его нельзя преодолеть регуляторным решением.

[10] Структурная несовместимость (*Institutional mismatch*) — третье из трёх структурных ограничений. Структурное несовпадение между устройством существующих институтов управления и тремя элементами реального правоприменения: полномочием на остановку, независимым доступом и последствиями за дезинформацию. Не нехватка ресурсов или политической воли, а архитектурное несовпадение между тем, для чего созданы институты, и тем, что требует управление ИИ.

[11] Остаточный управленческий ресурс (*Governance residual*) — авторская концепция, введённая в Эссе 12 серии. Совокупность механизмов ограничения — страхового андеррайтинг, судебные каналы, смежные регуляторные полномочия, — которые сохраняют реальный рычаг влияния даже после закрытия окна коррекции. Не синоним «остатков управления» в обычном смысле: это специфический технический термин, обозначающий частичный, асимметричный, но реальный потенциал ограничения.

[12] Архитектура управления (*Governance framework*) — совокупность норм, процедур, институтов и механизмов, определяющих, как принимаются решения, кто несёт ответственность и как реализуется надзор в области управления ИИ. В тексте также используется как «система управления» применительно к конкретным юрисдикциям или организациям.

[13] *Театр соответствия (Compliance theater) — практика формального соблюдения требований архитектуры управления при отсутствии реального эффекта ограничения. Аналог понятия «театр безопасности» (Security theater) в авиационном контексте: демонстрация мер без существенного снижения реальных рисков. Организации, расходующие остаточный управленческий ресурс на театр соответствия, поддерживают видимость контроля, истощая при этом последние реальные рычаги.*

[14] *Под термином «Глобальный Юг» (Global South) в данном анализе понимаются преимущественно развивающиеся и наименее развитые государства, которые в мировой технологической и институциональной архитектуре чаще выступают потребителями, а не разработчиками передовых систем ИИ и связанных с ними механизмов управления. Термин носит прежде всего социально-экономический и политический, а не строго географический характер, хотя многие такие государства действительно расположены в Южном полушарии или близко к нему.*

Приложение: Обзор серии и порядок цитирования

Таблица 2: Beyond Control — Эссе 1–12 с датами публикации и DOI

#	НАЗВАНИЕ / DOI	Опубликовано	ЦЕНТРАЛЬНЫЙ ТЕЗИС
1	Иллюзия контроля («The Illusion of Control») doi.org/10.5281/zenodo.19659250	12 фев 2026	Три механизма структурного сбоя — перформативный контроль, рассогласование стимулов, информационная асимметрия — прослежены от краха СССР 1991 года до современных систем управления ИИ
2	Ловушка прозрачности («The Transparency Trap») doi.org/10.5281/zenodo.19659603	17 фев 2026	Инструменты управления ИИ воспроизводят знакомую архитектуру: раскрытие информации без принудительной подотчётности. Прозрачность — ловушка, а не решение
3	Дилемма регулятора («The Regulator’s Dilemma») doi.org/10.5281/zenodo.19661370	23 фев 2026	Каждый регулятор встречает одно невозможное ограничение: понять технологию, успеть за ней, сохранить легитимность. Выберите два из трёх
4	Миф о согласовании («The Myth of Alignment») doi.org/10.5281/zenodo.19661522	3 мар 2026	Согласование — не техническая задача. Это вопрос власти: кто определяет ценности, кто их применяет и кто несёт последствия
5	Колониальный паттерн («The Colonial Pattern») doi.org/10.5281/zenodo.19661731	10 мар 2026	Управление ИИ воспроизводит паттерн старше самого ИИ: кто пишет правила — контролирует технологию.

#	НАЗВАНИЕ / DOI	Опубликовано	ЦЕНТРАЛЬНЫЙ ТЕЗИС
			Концентрация нормотворчества, извлечение без представительства
6	Паттерн замыкается («The Pattern Closes») <i>doi.org/10.5281/zenodo.19661816</i>	23 мар 2026	Механизмы, прослеженные в первых пяти эссе, стали операциональными одновременно в марте 2026 года в геополитике и отношениях ИИ-государство
7	Окно коррекции («The Correction Window») <i>doi.org/10.5281/zenodo.19661952</i>	30 мар 2026	Три элемента делают правоприменение реальным: последствия за дезинформацию, полномочие на остановку и независимая верификация с доступом. Все три отсутствуют
8	Трансфер агентности («The Agency Transfer») <i>doi.org/10.5281/zenodo.19664688</i>	6 апр 2026	Трансфер агентности — градиент с порогом, за которым откат становится операционно неосуществимым. Окно коррекции закрывается через тихую атрофию человеческих компетенций
9	Вопрос суверенитета («The Sovereignty Question») <i>doi.org/10.5281/zenodo.19664787</i>	13 апр 2026	Каждая архитектура правоприменения в управлении ИИ встречает одно структурное ограничение: суверенную волю государства, для которого технология стала элементом стратегической автономии
10	Вопрос инфраструктуры («The Infrastructure Question») <i>doi.org/10.5281/zenodo.19664876</i>	20 апр 2026	Физическая конфигурация вычислительного стека предопределяет пространство выбора для большинства юрисдикций до принятия любых суверенных решений. Предел материи
11	Институциональный разрыв («The Institutional Gap») <i>doi.org/10.5281/zenodo.19826978</i>	27 апр 2026	То, что требует реальное правоприменение, и то, что могут произвести существующие институты, — не одно и то же. Никакой объём регулирования не закрывает разрыв структурного происхождения
12	За пределами контроля: что происходит, когда окно коррекции закрывается («Beyond Control: What Happens When the Correction Window Closes») <i>doi.org/10.5281/zenodo.20067750</i>	4 мая 2026	Три ограничения — институциональное, суверенное, инфраструктурное — взаимодействуют мультипликативно. Остаточный управленческий ресурс: частичный, асимметричный, нескоординированный. Закрытие не объявляет о себе

Порядок цитирования данного синтеза

Khodjaev, O. (2026). "Beyond Control: Theory of Limits of AI Governance — An Analytical Synthesis." Version 1.0, May 2026. okhodjaev.com

DOI: doi.org/10.5281/zenodo.20120514

Глоссарий основных терминов

Окно коррекции (correction window)	Временной интервал между появлением системы и моментом, когда откат становится операционно неосуществимым.
Трансфер агентности (agency transfer)	Прогрессивная миграция полномочий по принятию решений от людей-операторов к автоматизированным системам через накопленную зависимость.
Остаточный управленческий ресурс (governance residual)	Совокупность механизмов ограничения, сохраняющих реальный операционный рычаг даже после того, как окно коррекции закрылось для системного управления передовыми системами ИИ.
Суверенный приоритет (sovereign override)	Структурный предел архитектур правоприменения, возникающий когда стратегический интерес государства в технологии вступает в противоречие с ограничениями архитектуры управления.
Инфраструктурная предопределённость (material predetermination)	Физическая конфигурация вычислительной, энергетической и производственной инфраструктуры, ограничивающая суверенные выборы в управлении ИИ до начала любых политических обсуждений.
Структурная несовместимость (institutional mismatch)	Категориальная несовместимость между существующими архитектурами управления и тремя элементами реального правоприменения: полномочием на остановку, независимым доступом и последствиями за дезинформацию.

Об авторе

Ойбек Ходжаев — основатель и генеральный директор INVEXI LLC, бывший заместитель председателя правления крупного узбекского коммерческого банка и бывший заместитель хокима Самаркандской области (2019–2022). Пишет об управлении ИИ и институциональных рисках на okhodjaev.com в рамках аналитической серии "Beyond Control — Theory of Limits of AI Governance".