31 июля 2026

ML — это не ИИ. Как непонимание терминов ломает инфраструктуру и бюджет

ML — это не то же самое, что генеративный ИИ. Точнее, в публичном поле эта разница почти стерлась: сегодня понятие ИИ фактически поглотило понятие ML. Но с точки зрения бизнеса эта унификация становится достаточно опасна. На фоне общего ажиотажа российский рынок наполнился идеями об «интеллектуализации всего», что при отсутствии глубокого технического понимания грозит превратиться в очередной технологический пузырь.

В данной статье мы не будем пытаться привести к общему знаменателю определение ИИ. Вместо этого сосредоточимся на разделении технологий по их назначению и принципам работы. Разграничение ML и ИИ критически важно, так как ошибка в понимании этих понятий ведет к применению не подходящих технологий, использованию неверной архитектуры систем и, как результат, неоправданно высоким затратам на внедрение инструментов.

В прошлой статье разбирали проблемы ИИзации в России, где ломается путь от идеи до эффекта.

Когда генерация подменяет анализ

Одна из наиболее частых ошибок при внедрении ИИ связана с неверным выбором инструмента под задачу. Компании формулируют аналитический

запрос — например, найти причины отклонений в расходах, выявить аномалии в операционных данных или определить источник финансовых потерь, — но вместо инструментов машинного обучения, которые применяются для подобных задач, выбирают генеративные платформы. Внешне это выглядит как современное ИИ-решение, однако по сути технология не соответствует природе задачи и методам для ее решения.

Генеративный ИИ хорошо работает там, где требуется создать новый текст, документ, отчет, резюме или ответ на запрос. Он также может быть эффективен для формирования ответов по определенным методологиям, если в него «вложить» методологические знания об этом процессе, путем настройки соответствующего алгоритма. Но если бизнесу нужно сформировать ответ и проанализировать уже существующие данные — обнаружить аномалии, найти закономерности, построить прогноз или классифицировать события, — такой подход становится ограниченным или вовсе неэффективным. Генеративная система может оформить результат для пользователя, представить его в удобной форме, но она не заменяет полноценный статистический и ML-анализ.

Именно здесь возникает ключевой риск: компания получает не отсутствие ответа, а отсутствие достоверной картины и верифицированных данных «под капотом» решения. Система может сгенерировать логичный на вид отчет, выделить формальные «проблемные зоны» и создать иллюзию проведенной аналитики. При этом реальные причины отклонений — ошибки в данных, сбои процессов, нестыковки в договорах, скрытые аномалии или повторяющиеся паттерны — остаются вне поля зрения. Аналитика на структурированных данных — это методы именно статистического анализа и ML.

Математическое и семантическое сходство — две разные логики работы с данными

Когда речь заходит о машинном обучении и искусственном интеллекте, часто упускают из виду принципиальную разницу между двумя способами, которыми модель «видит» данные: математическим сходством и семантическим сходством. Понимание этой грани — не академический вопрос, а практическая необходимость, от которой зависит, будут ли ваши ИИ-решения действительно предсказывать и анализировать, или просто интерпретировать данные на уровне контекста.

Математическое сходство — это, если грубо, количественная оценка. В этом контексте объекты сравниваются по их числовым характеристикам, а степень их «похожести» определяется через расчет разницы между определенными значениями, например, признаками объектов, выраженными в четких метрических оценках. Это работа с дистанцией: чем меньше числовая дельта, тем выше сходство.

Семантическое сходство переводит задачу из плоскости цифр в плоскость контекста. Здесь алгоритм оценивает близость объектов не по их собственным значениям, а по их «окружению». Если два разных объекта (например, слова или события) регулярно встречаются в одном и том же информационном поле, система интерпретирует их как семантически близкие.

Простой пример: при работе с массивами текстовой информации (например, при проверке договоров на наличие дубликатов или поиске заимствований) перед аналитиком встает фундаментальная задача — найти пересечения между двумя текстами. И здесь выбор между математическим и семантическим подходом определяет, будет ли результат полезным или просто декоративным.

Если ваша задача — найти точные текстовые совпадения, необходим строгий математический подход. Здесь считается «пересечение множеств»: сколько именно слов, символов или абзацев идентичны в обоих документах. Это

жесткая проверка на наличие дублирования: если в одном контракте указана сумма «10 000», а в другом «10 005» — математически это разные значения, и алгоритм должен зафиксировать это расхождение. Здесь нет места интерпретациям, есть только точный подсчет совпавших элементов.

Если же цель — найти «схожие по смыслу» фрагменты, в игру вступает семантика. Генеративный ИИ может сказать, что два текста «об одном и том же», даже если они написаны разными словами: один использует термин «оплата за услуги», а другой — «вознаграждение за выполненные работы». С точки зрения семантики — это пересечение по смыслу, но математически — по символам и строкам — это разные конструкции.

> Критическая ловушка заключается в том, что генеративный ИИ принципиально не предназначен для поиска и подтверждения точных математических пересечений.

Ловушка близости, или почему «похожесть» чисел может обмануть бизнес

Чтобы осознать масштаб риска, достаточно взглянуть на простую математическую аномалию. С точки зрения математического сходства числа 3 и 4 находятся на минимальном расстоянии друг от друга — разница между ними ничтожна, они «близки» по расчетам. Однако в рамках семантической логики (например, в задачах классификации уровней риска или критичности поломок) эти цифры могут представлять абсолютно разные категории: «низкий риск» и «критический сбой».

Именно здесь кроется главная опасность подмены ML-анализа генеративным ИИ. Когда задачи точных расчетов — в финансах, производстве, логистике —

пытаются решать с помощью систем, склонных к семантической интерпретации, результат оказывается ненадежным. Генеративный ИИ может «почувствовать» контекст и выдать красиво оформленный отчет, где цифры выглядят гармонично и логично, но при этом он полностью игнорирует математическую дистанцию между критическими показателями. В результате бизнес получает не точный прогноз, а правдоподобную галлюцинацию, которая подменяет жесткую математическую разницу между цифрами удобной семантической близостью.

Архитектурный тупик: когда структура данных определяет провал проекта

Фундаментальная причина, по которой подмена понятий «ML» и «ИИ» ведет к тяжелым последствиям для бизнеса, кроется не в маркетинге, а в глубоком архитектурном разрыве между способами работы со структурированными и неструктурированными данными. В основе любого технологического решения лежит вопрос: с чем именно мы работаем?

Машинное обучение (ML) работает прежде всего со структурированными данными. Классический ML — регрессии, деревья решений, случайные леса — опирается на таблицы, базы данных, логи и четкие признаки. Он работает с жесткой архитектурой: строка, столбец, числовое значение. Его задача — найти закономерность в упорядоченном массиве, где каждый элемент имеет свое математическое место. Это дисциплина, где ошибка в одном столбце (например, подмена даты на текст) может разрушить логику расчетов.

Генеративные ИИ-модели лучше подходят для работы с неструктурированными данными — текстами, изображениями, аудио. Их архитектура построена на извлечении семантически близких паттернов из неструктурированной информации. Они не видят таблицу как набор строго

заданных ячеек, а воспринимают ее как поток смыслов и контекстов.

Генеративный ИИ — это речевой аппарат, ML — это жесткий расчет. Проблемы начинаются в тот момент, когда функции этих двух систем смешиваются. Когда компания заменяет аналитический функционал (ML) генеративной моделью (ИИ), она фактически совершает архитектурную подмену. Мы берем речевой аппарат и заставляем его выполнять работу калькулятора. Использование инструментов для работы с неструктурированным смыслом (ИИ) там, где требуется точность структурных расчетов (ML) — это прямой путь к созданию систем, которые умеют красиво говорить, но совершенно не способны надежно считать.

За понятием «ИИ» и «ML» стоят абсолютно разные архитектурные подходы, и их подмена — это не просто ошибка выбора инструмента, а сбой в логике всей системы управления данными. ML-модели не могут работать в вакууме: им нужны единые, очищенные, структурированные данные. Хранилище данных критично для ML, поскольку обеспечивает структурированность, воспроизводимость и качество данных — того, что архитектура генеративного ИИ не может заменить, так как фокусируется на генерации контента, а не на управлении структурированными данными.

Последствия такой подмены напрямую отражаются на бюджете и управленческих решениях. Средства тратятся на внедрение технологически модного решения, которое не устраняет исходную проблему. Более того, ложная аналитическая уверенность может быть опаснее, чем отсутствие анализа: бизнес начинает принимать решения на основе красиво оформленных, но методологически слабых и необоснованных выводов.

Поэтому принципиально важно разделять задачи генерации и задачи анализа. Если цель — автоматизировать подготовку документов, отчетов или коммуникаций, генеративный ИИ может быть эффективным инструментом. Если же цель — выявить отклонения, прогнозировать риски, анализировать динамику или искать скрытые закономерности в данных, базовым выбором

должны быть ML-инструменты. ИИ не должен подменять аналитику там, где требуется не создание текста, а понимание данных.