

## Глава 2. Как стать создателем ценности с помощью ИИ

Это продолжение перевода книги Томас и др. Создание бизнес-ценности с генеративным ИИ. В предыдущей главе мы предложили набор практических принципов, которые могут сразу повысить шансы на успех в любой инициативе, связанной с ИИ. Все это основано на нашем коллективном опыте — от тысяч клиентских проектов до участия в телешоу *60 минут*, взаимодействия с Белым домом, НАТО, топ-менеджментом крупнейших компаний и даже с Ватиканом! (В Ватикане хранятся бесценные артефакты в азотных хранилищах, и мы — ну, точнее, более широкая команда IBM — помогли сделать эти артефакты доступными для исследователей, чтобы масштабировать знания и сохранить историю. Не можем раскрыть подробности этой сделки, но мы уверены в посмертной благодарности.)

| Традиционный ИИ<br>Обучение с учителем          | Новый подход ИИ<br>Самообучение                                   |
|---|---|
| Зависит от человеческого труда                  | Основано на вычислительных мощностях                              |
| Требует тщательной разметки данных              | Требует минимальной разметки                                      |
| Долгий, сложный, утомительный и дорогой процесс | Быстрое, автоматизированное и эффективное благодаря трансформерам |

Рис. 2.1. Сравнение энергии активации для начала работы при обучении с учителем и самообучении

[Предыдущая глава](#)   [Содержание](#)   [Следующая глава](#)

Вы также узнали, почему текущий момент сравним с «моментом Netscape» — волной изменений, которая захлестнет и вашу личную, и профессиональную жизнь. Как когда-то электричество казалось волшебством (хотя это было не так), так и ИИ не является магией. Мы подтолкнули вас (ладно, грубо подтолкнули обеими руками) к переходу от мышления +AI к мышлению AI+, а также предложили адаптированную под текущую реальность *Лестницу ИИ*, которая поможет добиться успеха. Наконец, мы дали вам прикладные подходы: как классифицировать бюджеты, выбирать кейсы и определять ожидаемые эффекты — будь то упрощение процессов (shift left) или выход на рынок (shift right).

Мы считаем, что и первая, и эта глава критичны: обе помогают сфокусироваться на деталях, которые действительно имеют значение на пути внедрения ИИ. Почему? Потому что именно детали определяют качество, выделяют вас на фоне других и формируют доверие. Позвольте привести аналогию со Статуей Свободы. Она возвышается в гавани Нью-Йорка, окутанная зеленой патиной — продуктом естественной химической реакции меди с окружающей средой. Но когда-то она сияла медным блеском — именно такой ее впервые увидели иммигранты, входившие в порт. Обратите внимание на ее прическу: на старых фото крупным планом видна тончайшая работа — идеальные локоны, детально проработанная коса. Скульптор Фредерик Огюст Бартольди создал ее за десять лет до первого полета человека — у него не было оснований полагать, что кто-то когда-либо разглядит прическу с высоты. Но детали имели значение, потому что скульптура была его ремеслом, и его репутация зависела от качества исполнения.

Причем здесь ИИ? Решения, которые вы примете в ближайшие годы, возможно, никто никогда не увидит в отрыве от других — но детали будут иметь значение. Они станут отражением вас, вашей команды и вашей компании. Помните об этом.

В этой главе мы хотим познакомить вас, пожалуй, с самым важным «пунктом назначения» на вашем ИИ-маршруте: с ролью **создателя ценности с помощью ИИ (AI Value Creator)**. Помните, что мы все еще находимся на стороне бизнеса, поэтому хотя мы и будем давать технологические пояснения по крупным языковым моделям (LLM) чуть позже, сейчас мы хотим расширить ваш управленческий инструментарий, чтобы вы могли извлекать максимум пользы из ИИ — в отличие от тех, кто эту книгу не читал.

## *ИИ сквозь десятилетия: «ускоренная съемка» развития*

Термин искусственный интеллект появился в 1956 году, и с тех пор сменилось несколько поколений технологий ИИ — каждое со своими прорывами и разочарованиями. Многие утверждают, что разочарований было больше, чем вдохновляющих достижений. Это привело к так называемым «зимам ИИ» — периодам спада интереса, из которых ИИ возвращался после ключевых технологических прорывов. Впрочем, если посмотреть на историю других изобретений — например, электричества — становится ясно, что путь к практическим успехам в ИИ неизбежно шел через массовые эксперименты. Многие попытки оказывались неудачными, но именно неудачи помогали выявить и решить проблемы, стоявшие на пути прогресса.

Идея о том, что машины смогут мыслить как люди, существует с конца XIX века. Однако по-настоящему она оформилась в 1950 году благодаря работе Алана Тьюринга [Вычислительные машины и разум](#). Историки называют его отцом искусственного интеллекта. В статье он рассуждал о том, как компьютеры смогут играть в шахматы, превзойдут людей в этом и научатся понимать естественный язык. Он предположил, что машины однажды научатся мыслить.

За свою карьеру в IBM авторы книги стали свидетелями (а иногда и участниками) воплощения многих идей Тьюринга: это и Deep Blue, победивший чемпиона мира по шахматам, и Watson, выигравший в Jeopardy!, и системы, играющие в го или участвующие в дебатах. Но Тьюринг был лишь началом.

Если работа Тьюринга была искрой, то своего рода «большой взрыв» произошел шесть лет спустя — в 1956 году — во время летнего семинара в Дартмутском колледже. Тогда группа молодых ученых вместе с исследователями из Bell Labs и IBM организовала научную встречу, чтобы обсудить перспективы искусственного интеллекта. Именно на этом семинаре впервые прозвучал сам термин «искусственный интеллект», и именно он считается официальной точкой отсчета области как самостоятельной дисциплины.

На том семинаре были определены основные задачи, над которыми исследователи работают до сих пор: нейросети, обучение без вмешательства человека, машинное творчество и многое другое.

Для понимания масштаба: тот же 1956 год — это год, когда за изобретение транзистора была вручена Нобелевская премия. Сегодня мы можем уместить более 100 миллиардов транзисторов в одном графическом процессоре (GPU) и объединять множество таких GPU, чтобы создавать вычислительные кластеры, необходимые для генеративного ИИ. На протяжении десятилетий теории и методы ИИ развивались параллельно с аппаратным обеспечением, и сегодня они сошлись в одной точке, делая ИИ реальным, практичным и масштабируемым.

Но ключевое, на чем настаивают авторы: дело не только в мощных чипах и хитрых алгоритмах. Самым важным элементом в генеративном ИИ — особенно если говорить о ценности для бизнеса — являются данные. Без них невозможно реализовать потенциал таких систем. Поэтому *оборудование, алгоритмы и данные* — это три равнозначные опоры, на которых держится весь ИИ.

### *Коротко о базовых моделях*

В мире генеративного ИИ (GenAI) вы часто будете слышать, что LLM — это движущая сила GenAI. Но что это такое? На базовом уровне LLM (large language model) — это способ представления языка в многомерном пространстве с огромным числом параметров. Эти представления формируются путем обучения модели на гигантских объемах текстов.

Небольшое отступление для разъяснения терминов, которые будут регулярно встречаться и в этой книге, и в повседневной деловой коммуникации.

**LLM.** Это наиболее распространенный тип моделей в GenAI. Во всех случаях, когда в книге идет речь о GenAI, по умолчанию подразумеваются именно LLM, если явно не указано иное. К таким моделям относятся: OpenAI ChatGPT, Google Gemini, Meta Llama, IBM Granite, DeepSeek и Mistral AI — все они представляют собой большие языковые модели.

Кроме LLM существуют и другие типы моделей. Например, диффузионные модели, которые создают изображения, видео или другие виды контента на основе текстовых запросов. Они работают по принципу добавления шума к данным (например, к изображению кота), а затем обучения модели «обратно» восстанавливать из шума исходное изображение. В терминах ИИ этот процесс проходит

через множество циклов (эпох). ИИ по сути учится восстанавливать изображение из зашумленной версии.

**Параметры.** LLM часто обозначают по количеству параметров. Например, Llama-3-70B означает, что модель содержит 70 миллиардов параметров. В этом контексте параметр — это числовое значение, часть внутренней структуры модели, которое отвечает за хранение знаний, полученных при обучении. В книге также упоминаются веса модели — это практически то же самое. Множество этих параметров (или весов) — это и есть «мозг» модели. Как правило, чем больше параметров, тем больше задач модель может выполнять. Но больше — не всегда лучше. Размер модели должен соответствовать задаче. Если, например, модель используется для отправки писем с напоминаниями о задолженности, нет смысла использовать огромную модель, умеющую писать в стиле Джоуи Трибиани из Friends: «Как дела?» — и сразу сумма долга. Или в стиле Майкла Скотта из The Office.

### *Пространство высокой размерности*

Это может показаться сложным, но мы постараемся объяснить просто. Представьте песню и опишите ее в трех измерениях (3D). Легко, правда? Допустим, вы выбрали «Shake It Off» Тейлор Свифт. (Будем честны — неважно, любите вы ее или нет, вы все равно знаете все слова, так что даже не...) Мы бы описали эту песню как {поп, вдохновляющая, жизнеутверждающая}, и, конечно, как вы узнали в Главе 1, для ИИ все это — просто числа.

Теперь попробуйте описать эту песню в 10 измерениях (10D). У нас получилось: {поп, запоминающаяся, вдохновляющая, дерзкая, веселая, гимническая, жизнеутверждающая, радостная, энергичная, игривая}. А теперь попытайтесь визуализировать эти 10 измерений на графике — у вас получится пустое пространство. (Надеемся, вы оцените иронию.) Если у вас еще не разболелась голова, попробуйте описать эту песню, используя сто измерений, а затем тысячу. Проще говоря, когда специалисты по алгоритмам говорят о данных в высокоразмерном пространстве, они имеют в виду ситуации, где измерений так много, что их невозможно представить.

Люди не способны мыслить в высокоразмерных пространствах, но ИИ может оперировать огромным количеством измерений. Например, каждая песня в Spotify кодируется (то есть представляется в виде чисел) сотнями, если не тысячами измерений, которые численно описывают композицию. Данные в высокоразмерном пространстве открывают для ИИ множество возможностей.

Рассмотрим рекомендательную систему Spotify. Плейлист пользователя — это как предложение, состоящее из тысяч измерений, отражающих его музыкальные предпочтения. Допустим, в плейлисте сильно выражены опера, классическая музыка и поп. Spotify может порекомендовать «Bohemian Rhapsody» Queen (отличный выбор, если уж на то пошло) из-за оперных элементов в песне. Это может привести к дальнейшему увлечению опероподобными композициями — и внезапно вы слушаете «B.Y.O.B.» System of a Down. Почему? Где-то среди тысяч измерений, описывающих эти песни, наверняка есть параметры, указывающие на оперность, классические нотки или повествовательность. («Bohemian Rhapsody», кстати, рассказывает о молодом человеке, который убил кого-то и продал душу дьяволу.) Все это возможно потому, что если для нас песня имеет всего несколько характеристик, то для ИИ Spotify — тысячи. И пока вы не можете удержать в голове тысячи измерений, ИИ может — и благодаря этому вы наслаждаетесь идеально подобранным плейлистом, гуляя, скажем, по лондонскому рынку Камден (какая ирония).

С этой точки зрения, большая часть истории вычислений — это поиск новых способов представления данных и извлечения из них пользы. Долгое время мы хранили данные в таблицах. Например, сотрудники или клиенты были строками в базе данных, а их атрибуты — столбцами. Это отлично подходит для обработки транзакций (OLTP) или выписывания чеков.

Затем мир начал представлять данные в виде графов, что помогло обнаруживать и анализировать связи между точками данных так, как раньше было невозможно. Например, этот человек, бизнес или место связаны с другими людьми, бизнесами или местами. Данные, представленные таким образом, начинают раскрывать закономерности. Компании используют графы для анализа социальных сетей или выявления подозрительных покупок, чтобы обнаружить мошенничество с кредитными картами. Эта технология сочетает множество подходов к анализу данных с использованием различных хранилищ (включая графовые базы данных). Именно так работает функция «Люди, которых вы можете знать» (PYMK) в Facebook — и это лишь один пример.

Сегодня с LLM мы работаем с огромными массивами данных, представленными в нейронных сетях, которые (очень приблизительно) имитируют мозг. Эти сети состоят из множества слоев с миллиардами или даже триллионами параметров - и внезапно появляется возможность делать удивительные вещи. Можно обнаружить настолько детальные закономерности, что предсказания становятся гораздо более уверенными. Например, можно предсказать, что это слово, скорее всего, связано со следующим словом, а эти два слова, вероятно, будут сопровождаться определенным третьим словом - и так можно постепенно строить, перепроверять и снова предсказывать, пока не будет создано или сгенерировано что-то новое. Отсюда и термин "генеративный ИИ".

GenAI - это способность анализировать данные, выявлять взаимосвязи и с достаточной уверенностью предсказывать последовательности, чтобы создавать или генерировать то, чего раньше не существовало. Текст, изображения, видео, звуки и практически любые типы данных могут быть представлены в модели.

Раньше мы могли делать ограниченные версии всего этого с помощью глубокого обучения, которое само по себе было важной вехой в развитии ИИ. С глубоким обучением мы начали представлять огромные объемы данных с помощью очень больших нейронных сетей со множеством слоев, но обучение требовало аннотированных данных, которые люди должны были вручную маркировать - например, отмечать на изображении "кошка" или "собака". Это называется обучением с учителем. Проблема была в том, что такое обучение, как показано на рис. 2.1, оказалось дорогим, трудоемким и длительным процессом, поэтому только крупные институты занимались этой работой и только для конкретных задач. Если вы хотели, чтобы ИИ мог суммировать и переводить текст, вам нужно было вручную разметить два очень больших набора данных (об этом чуть позже).

Около 2017 года появился новый подход, основанный на архитектуре под названием трансформеры (мы кратко рассмотрим их в Главе 9). С этим подходом ИИ смог выполнять новый вид беспрепятственного обучения, называемый самообучением, при котором языковая модель могла обучаться на больших объемах немаркированных данных, скрывая определенные части текста (слова, предложения и т.д.) и прося модель заполнить пропуски (на жаргоне ИИ это называется маскированием). Например, если мы скажем "Да пребудет с тобой", вы, скорее всего, догадаетесь, что следующие три слова - "Сила" из "Звездных войн". Хотя это упрощение, этот удивительный процесс, выполняемый в больших масштабах, приводит к мощным представлениям данных, которые сегодня мы называем LLM.

Именно здесь произошло нечто поистине волшебное. Исследователи обнаружили, что вместо создания моделей ИИ, пригодных только для узких случаев использования и областей знаний (например, трудоемкого создания одного набора данных для суммирования и другого для перевода), они могут получить ИИ с более широкой применимостью. По сути, эти LLM можно было обучать на огромных объемах интернет-данных (сегодня самые популярные LLM - это просто сильно сжатые представления всего, что есть в интернете, что имеет как плюсы, так и минусы), и таким образом они приобретали человеко-подобные возможности естественного языка.

Самообучение в больших масштабах в сочетании с огромными объемами данных и вычислительных мощностей подарило миру ИИ, который является *универсальным* и *адаптивным*. Мы определяем эти термины следующим образом:

**Универсальность** (Generalizable). ИИ способен выполнять широкий спектр задач в различных областях, часто без необходимости специальной настройки под конкретную задачу. Другими словами, одна и та же LLM, которая классифицирует тональность текста, может извлекать имена и места из текста (распознавание именованных сущностей, NER), а также переводить, суммировать и многое другое.

**Адаптивность** (Adaptable). Это означает, что ИИ может не только выполнять несколько задач, но и справляться с вариантами использования, для которых изначально не обучался. Адаптивный ИИ также обладает эмерджентными свойствами — способностями, которые не были явно запрограммированы и возникают неожиданно. Например, LLM может разгадывать загадки или решать логические головоломки, на которых она никогда не обучалась, просто распознавая закономерности. Суть в том, что возможность использовать одну и ту же модель для множества задач и открывать в ней новые способности — это мощный инструмент (хотя вам все равно нужно направлять ее, чтобы стать создателем ценности с помощью ИИ; об этом чуть позже).

За последнее десятилетие произошел взрывной рост применения ИИ. (Мы уверены, что вы использовали многие из этих технологий, даже не осознавая этого. Пользовались Siri или Alexa? Меняли серое небо на солнечное, чтобы создать идеальное фото? Использовали приложение для перевода?) За это время ИИ превратился из чисто академического направления в мощную силу, которая влияет на множество отраслей и ежедневно затрагивает жизни миллиардов людей.

В последние годы нам удалось создать ИИ-системы, способные обучаться на миллионах примеров, чтобы лучше понимать наш мир и находить новые решения сложных проблем. Эти масштабные модели привели к разработке систем, которые понимают нас, когда мы говорим или пишем. Сюда входят технологии обработки (NLP) и понимания (NLU) естественного языка, которые мы используем каждый день — от цифровых ассистентов до программ преобразования речи в текст. Другие системы, обученные, например, на всех работах известных художников или учебниках по химии, позволили создать генеративные модели, способные создавать новые произведения искусства в стиле этих художников или разрабатывать новые химические соединения на основе истории исследований.

Хотя сегодня многие новые ИИ-системы помогают решать самые разные реальные проблемы, до появления GenAI создание и развертывание ИИ для каждой новой системы традиционными методами требовало значительных временных и ресурсных затрат. Для каждого нового приложения необходимо было иметь большой, хорошо размеченный набор данных под конкретную задачу. Если такого набора не существовало, люди тратили сотни или тысячи часов (а то и больше) на поиск и разметку изображений, текстов или графиков для обучающих и проверочных данных.

Что это значит для вас? Вы можете взять большую предобученную LLM (если используете ее для бизнеса, важно начать с надежной модели) и добавить свои корпоративные знания, чтобы усилить ее для решения ваших конкретных задач с вашими данными. (Подробнее о плюсах, минусах и подводных камнях этого подхода — в Главе 8.)

Если вы сейчас чувствуете разочарование, потому что ваш бизнес относится к тем, кто потратил огромное количество времени на сбор и разметку данных для ИИ-проектов, которые в итоге провалились из-за недостаточного объема размеченных данных (именно так часто происходило с традиционным ИИ), не отчаивайтесь! Эта работа не пропала даром в эпоху GenAI, потому что именно эти специализированные отраслевые данные вы теперь сможете использовать для настройки LLM под нужды вашего бизнеса. Это именно то, что нужно, чтобы стать создателем ценности с помощью ИИ. Более того, вы буквально сможете взять те самые «провальные» ИИ-проекты двухлетней давности и предстать героем в глазах руководства, предложив, как адаптировать выбранную LLM для ваших задач.

Как это возможно? Во-первых, современные LLM почти не содержат корпоративных данных (всего около 1%), не говоря уже о ваших уникальных данных. В Главе 1 мы объясняли, что ваши данные — это конкурентное преимущество, и теперь пришло время заставить их работать.

Проще говоря, когда вы объединяете представления данных LLM с вашими размеченными данными (которых теперь нужно гораздо меньше), вы получаете решение, идеально подходящее для вашего бизнеса. Представьте, что вы знаете испанский и теперь учите французский. У вас уже есть база: вы понимаете, как спрягать глаголы или строить предложения. Так же и с LLM: если у вас есть основа, адаптация к новым задачам становится проще. В Главе 8 вы узнаете о новом открытом подходе (InstructLab), который позволяет легко интегрировать ваши данные в приватную LLM компании, не раскрывая их миру, — и это гарантированно улучшит ваши результаты.

Сейчас LLM чаще всего ассоциируются с обработкой естественного языка. Но возникает вопрос: что такое язык? Сигналы промышленного оборудования — это язык; программирование — это язык для общения с машинами; клики пользователя на сайте, химические формулы, даже моделирование вкусов и запахов — все это языки. Если присмотреться, все вокруг можно рассматривать как язык, а значит, его можно изучить, расшифровать и понять.

Главный вывод: ИИ можно специализировать для решения самых разных задач, повышающих продуктивность в любом «языке». Это означает, что ИИ может применяться горизонтально во всех сферах вашего бизнеса: HR-процессы, обслуживание клиентов, кибербезопасность, написание кода, модернизация приложений и многое другое, о чем мы расскажем в Главе 4.

## Эволюция больших языковых моделей и сравнение обучения с учителем и самообучения

Большие языковые модели устроены иначе, чем традиционный ИИ. Они обучаются методом самообучения, что исключает необходимость ручной разметки огромных массивов данных. По сути, модель обучается, просто "читая" колоссальные объемы информации (например, текста), и в результате получается универсальная система с почти человеческими языковыми способностями. ИИ использует математические модели для представления взаимосвязей в данных (например, между словами). Если дать модели несколько слов из подсказки, она может математически предсказать наиболее вероятное продолжение — как в примере с фразой из "Звездных войн", который мы привели ранее.

Два ключевых преимущества GenAI, которые нас особенно вдохновляют:

- Невероятная скорость создания решений благодаря факторам, указанным на рис. 2.1.
- Универсальность и адаптивность моделей, о чем мы говорили в предыдущем разделе.

Чтобы по-настоящему оценить, как GenAI сокращает время реализации ИИ-проектов, нужно выйти за рамки разметки данных и сравнить этот подход с традиционными методами внедрения ИИ.

Многие из вас, кто давно работает с ИИ, возможно, заметили, что сценарии использования из эпохи традиционного ИИ теперь повторяются в эру GenAI — и вы абсолютно правы. Однако стоит отметить, что хотя начальные применения GenAI могут казаться знакомыми, появляются и совершенно новые возможности, особенно с развитием агентного ИИ.

В последнее десятилетие с появлением глубокого обучения мир доказал, что можно достичь впечатляющей точности в конкретных задачах, если собрать достаточно данных, разметить их, обучить модели и внедрить их. Эта традиционная методология показана на рисунке 2-2.

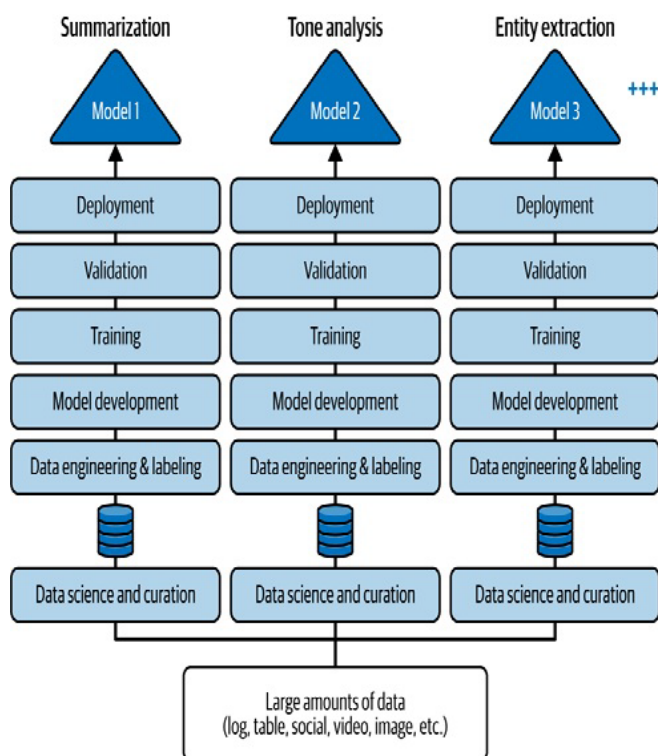


Рис. 2.2. Традиционный подход к созданию ИИ: формирование множества команд data science для параллельной работы над разными проектами

Каждая модель создается для конкретного сценария. В данном примере это суммирование, анализ тональности и извлечение сущностей. При традиционном подходе ваша компания создавала бы отдельную команду для каждой задачи, и каждая команда разрабатывала бы отдельную модель. Все эти команды проходили бы один и тот же трудоемкий процесс: отбор и подготовка данных, разметка, разработка модели, обучение, валидация — часто дублируя одни и те же данные!

GenAI кардинально меняет эту парадигму. Вместо множества узкоспециализированных моделей теперь можно использовать одну универсальную LLM, адаптируя ее под разные задачи с гораздо

меньшими затратами. Это как заменить целый оркестр отдельных музыкантов одним универсальным синтезатором, который может воспроизводить любые инструменты по требованию.

Переход от традиционного ИИ к GenAI напоминает эволюцию от специализированных фабрик прошлого к гибким автоматизированным производствам сегодня — где одна перенастраиваемая линия может выпускать разные продукты без полной перестройки. Именно эта гибкость открывает путь к настоящей демократизации ИИ, делая его доступным не только для технологических гигантов, но и для обычных бизнесов.

Разные команды, собирающие данные и подготавливающие их для своих конкретных задач, проходя одни и те же этапы, что и другие команды — это процесс, который можно описать только как долгий, сложный, утомительный и дорогостоящий. По сути, масштабирование ИИ в вашей компании сводилось к ответу на вопросы: сколько команд data science вы можете собрать и сколько проектов эти команды смогут реализовать?

Теперь сравните новый подход к ИИ (левая часть рис. 2.3) с традиционным путем (правая часть рисунка). Как видите, вместо создания отдельной модели ИИ для каждой конкретной задачи, вы берете LLM, которая, скорее всего, уже обучена кем-то другим (например, IBM, Google, DeepSeek, OpenAI или Anthropic; честно говоря, немногие компании будут создавать свои собственные — вместо этого они будут адаптировать существующие), и настраиваете ее для множества различных задач. Обратите также внимание, как одна LLM поддерживает три варианта использования:

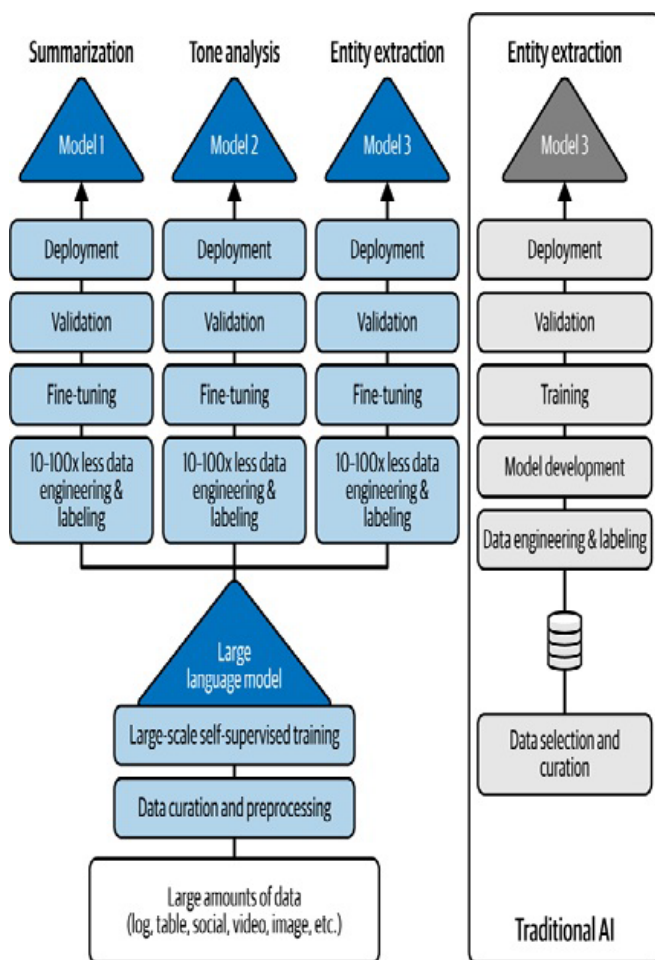


Рис. 2.3. GenAI масштабирует ИИ, сокращая требования к навыкам, данным, времени, администрированию и первоначальным затратам

Благодаря универсальности LLM компании теперь могут использовать одну и ту же модель для реализации нескольких бизнес-сценариев. С традиционным ИИ это было невозможно. Мы настоятельно рекомендуем вам запомнить рис. 2.3, так как он иллюстрирует, почему LLM становятся ключевым компонентом нового рабочего процесса ИИ. Современный ИИ требует значительных усилий для создания базовой модели (универсальной LLM), но затем позволяет получить эффект масштаба от этих инвестиций. Создание собственной LLM — сложная задача, поэтому мы уверены, что

большинство из вас выберет готовую модель и адаптирует ее под свои бизнес-потребности (как это сделать, мы расскажем далее в книге).

Надеемся, вы уловили суть этого методологического сдвига, потому что следующая волна ИИ заменит специализированные модели, которые доминировали до сих пор, на LLM в качестве ядра. Эти модели обучаются на широком наборе данных, которые можно использовать для разных задач, а с их способностью к самостоятельному поиску решений для достижения целей, агентный ИИ пойдет по тому же пути.

Вот главный вывод. Универсальность LLM заключается в том, что они, как следует из названия, могут стать основой для множества ИИ-приложений и агентных систем. Благодаря самообучению и переносу знаний эти ИИ-модели могут применять полученные навыки к новым типам задач.

Проще всего понять перенос знаний на классическом примере компьютерного зрения, где ИИ учится распознавать кошек. (Снова ИИ и кошки идут лапа об лапу — словно какой-то любитель кошек решил, что его глубокому обучению нужно глубокое мурлыканье.) Если вы обучаете ИИ определять кошку, он начинает с распознавания границ и форм, постепенно наращивая слои нейронной сети, чтобы идентифицировать кошку. На базовом уровне такой ИИ научится находить треугольники (комбинации граней).

Если подумать о кошке, треугольники формируют ее уши, нос и другие части. Как только ИИ научится находить треугольники, он сможет выявлять другие особенности кошки, используя все больше слоев нейронной сети, пока окончательно не определит объект как кошку. Теперь представьте, что вы хотите научить ИИ распознавать парусник. Модель, обученная находить парусники, начнет с того же — поиска границ и форм. Вы можете взять слои ИИ, которые уже умеют распознавать треугольники, и применить их для парусников. То же можно сделать для тысяч слоев — вот вам и перенос знаний. Независимо от того, распознает ли ИИ кошку или парусник, умение находить треугольники остается критически важным.

Большинство из нас на бытовом уровне сталкивается с универсальностью LLM, поддерживающих множество задач. Например, если вы научились водить автомобиль, эти навыки можно перенести на управление другими машинами. Конечно, придется привыкнуть к нюансам (например, где включаются дворники), а в некоторых случаях возникнут сложности (попробуйте сесть за руль с механической коробкой, если до этого ездили только на автомате), но базовые навыки останутся. Сегодня никто не создает сверточные нейронные сети (CNN) или vision transformer (ViT) для компьютерного зрения без переноса знаний — это как главный чит-код в этой области!

Вывод прост: вместо создания отдельной модели для каждой задачи можно обучить одну модель и адаптировать ее для множества различных применений. Это означает, что компании теперь могут перейти от парадигмы «одна задача — одна модель» к подходу «одна модель — много задач». Например, ваш ИТ-чатбот, HR-портал и приложение для написания маркетинговых писем могут использовать одну базовую модель.

Как показано на рис. 2.3, работа все же требуется! Хотя объемы инженерии данных и разметки теперь минимальны, вам все равно нужно будет настроить модель под специфику вашего бизнеса: бренд, стиль, нормы и т. д. Для этого есть множество методов, включая настройку промптов (prompt tuning), разработку промптов (prompt engineering), тонкую настройку с помощью PEFT (Parameter-Efficient Fine-Tuning) и метод InstructLab. Подробнее об этом — в Главе 8, но благодаря LLM подготовительная работа значительно сократилась.

Главное здесь — не мощность модели с миллиардами или даже триллионами параметров, а то, что LLM позволяют масштабировать ИИ-инициативы с меньшими затратами времени, данных, денег и администрирования. Например, в IBM традиционным способом поддержка 12 языков заняла 7 лет, но с переходом на GenAI их число выросло до 25 всего за год.

### *Создание ценности с помощью ИИ должно быть вашей конечной целью*

Когда соединяются кислород, тепло и топливо, мы получаем огонь. Это просто, фундаментально и стало ключом к прогрессу человечества. Подумайте: огонь дал свет, тепло и защиту, позволил нашим предкам осваивать новые климатические зоны и употреблять новую пищу. Гончарное дело, металлургия, химия, скоростной транспорт и многие другие технологии начались с огня. Но представьте, если бы огонь был проприетарной технологией? Если бы знания о его создании не



распространялись и существовало бы лишь несколько хранителей этого секрета? Где бы мы сейчас были?

Вспомните, что мы говорили в предисловии: мы находимся в моменте "подъема, сдвига, разлома или обрыва" с GenAI, и особенно с агентами, которые будут формировать наше общество в ближайшие поколения. Этот раздел (и остальная часть книги) покажет вам, как стать создателем своего "ИИ-огня", взять под контроль свою ИИ-судьбу и почему так важно воспринимать себя как создателя ценности с помощью ИИ, а не просто пользователя. Наконец, мы подробно объясним, почему будущее ИИ нуждается в экосистеме открытых инноваций.

### *Как вы потребляете ИИ: создатель ценности или пользователь?*

Существует три режима использования ИИ:

1. Он встроен в программное обеспечение
2. Вы используете чужую модель
3. Вы используете платформу ИИ

### **Пользователь ИИ: "встряхни" (встроенный) и "испеки" (в продукте) ИИ**

Первый способ использования ИИ — когда он "встроен" в готовое программное обеспечение. В этом подходе вендор создает ИИ, а вы его используете. (Мы предполагаем, что вы работаете только с настоящим ИИ в продуктах, а не с "фейковым ИИ", поскольку сегодня все заявляют о наличии ИИ в своих продуктах.) Будь то помощник в написании текстов (вроде Grammarly или Jasper), помогающий найти правильный тон для письма, или ПО для редактирования изображений (например, Adobe Photoshop или Topaz Photo AI), автоматически улучшающее качество ваших фото и видео — с этим подходом вы, как пользователь ИИ, получаете доступ к мощным функциям, повышающим вашу продуктивность. Кто откажется от такого?

Но есть нюанс! Вы и все остальные получаете доступ к одной и той же "магии", а значит, хотя такой ИИ помогает вам работать быстрее и качественнее (это хорошо), он делает то же самое для любого, кто потратит время на освоение этого ПО. Другими словами, эти возможности ИИ не становятся конкурентным преимуществом — они лишь устанавливают новую, более высокую планку для всех, включая ваших конкурентов.

### **Пользователь ИИ: не упави, когда делаешь сервисный запрос (еще большее "но")**

Второй способ использования ИИ — когда вы отправляете запросы чужой модели, либо через чат-интерфейс, либо через API. Проще говоря, разрабатывая кастомные ИИ-приложения для бизнеса, вы можете обращаться к GenAI-сервису другой компании, использовать их модели и получать результаты. Это тоже рабочий подход.

Правда в том, что почти все мы так использовали GenAI, оставаясь обычными пользователями ИИ. Но задумайтесь: вы ограничены возможностью просто запрашивать чужую модель (не вашу), не контролируете ни саму модель, ни данные, на которых она обучалась, и в большинстве случаев понятия не имеете, какие именно данные легли в ее основу.

В зависимости от того, насколько искусно вы используете модель, можно создать некоторое отличие от конкурентов. Но есть серьезные оговорки, особенно если вы стремитесь стать создателем ценности с помощью ИИ.

Во-первых, как и в случае с ПО, эти модели и сервисы доступны всем. Так есть ли реальное отличие? Да, возможно, вы формулируете промты лучше других. Но вы все равно используете ту же модель, что и все.

Есть и более тревожный аспект: когда ваше приложение делает запрос, оно подключается к чему-то непрозрачному. Вы не знаете, что происходит на другом конце: как модель обрабатывает ваши данные (учится на них, хранит или просто анализирует), каково происхождение данных, на которых обучена LLM. В зависимости от сценария использования это должно вызывать беспокойство — ведь ваш бизнес несет ответственность за конечный результат (социальную или, все чаще, юридическую — об этом в Главе 5). Для бизнес-использования ИИ (в отличие от личного) это особенно критично.

Второй момент для размышления: кто в долгосрочной перспективе получает основную ценность? История знает множество экстрактивных бизнес-моделей — если вы в соцсетях, вы часть одной из них. Как мы говорим: если вы не платите за продукт, то, скорее всего, вы и есть продукт. Даже платя за сервис, вы получаете ценность (иначе зачем платить?), но компания-провайдер тоже извлекает выгоду

из вашего использования и данных, накапливая все больше со временем. Мы не будем называть имен, но примеров компаний (включая платные сервисы), получающих стратегическую выгоду от ваших данных, множество. Ирония в том, что именно так создавались многие LLM — путем сбора данных из интернета, защищенных авторским правом или нет.

Главный вопрос: если вы, как пользователь ИИ, делаете запросы к чужому сервису, чья ценность растет быстрее — ваша или их? (Подсказка: посмотрите на акции и мультипликаторы оценки некоторых компаний.) Баланс в этих отношениях нарушен, что может иметь долгосрочные последствия для вашего бизнеса, экономики и технологического прогресса.

Финальное "но": действительно ли мы, как общество, хотим зависеть от горстки "хранителей" ИИ-огня? Выгодно ли это вашему бизнесу и акционерам? Мы считаем, что нет.

### Огненный стартапер: начало создания ценности с помощью ИИ

Третья модель потребления ИИ — это модель платформы, которая является самой всеобъемлющей. Это то, как вы становитесь стартапером ИИ. Это не значит, что вы действуете в одиночку или заново изобретаете ИИ с нуля. Вы не тратите годы и миллионы на создание собственных LLM. Конечно, вы можете сделать это с платформой, но это будет в очень малом числе случаев. С платформой для создания ценности с помощью ИИ у вас есть все элементы и компоненты (данные, управление, LLM), чтобы создать свои собственные ИИ-решения. У вас есть доступ к огромному числу моделей GenAI (как с открытым исходным кодом, так и платных), или вы можете загрузить свои собственные модели на платформу. У вас есть инструменты для улучшения и настройки моделей, чтобы интегрировать ваши собственные знания о бизнесе без опасений по поводу передачи ваших данных в Интернет.

Вы можете точно настраивать модели, подсказки, адаптировать их с помощью InstructLab — любые методы, которые мы описываем в главе 8. Такой подход позволяет вам создавать и накапливать ценность, которая уникальна для вашего бизнеса.

Отличный пример создания ценности с помощью ИИ — это L'Oréal, одна из ведущих мировых косметических компаний. Представьте себе корпус данных о формулах, материалах и предпочтениях, которые L'Oréal накопила, приближаясь к своему 120-летию. По сути, L'Oréal обладает данными, которые определяют язык макияжа. Она хочет стать создателем ценности с помощью ИИ, поэтому решила создать частную ИИ-модель (в сотрудничестве с IBM), чтобы ускорить разработку новых продуктов, переработку существующих косметических средств и оптимизировать процессы масштабирования производства. Если бы L'Oréal была просто пользователем ИИ, она бы отдала эти данные, но вместо этого рассматривает их как конкурентное преимущество и решила использовать их для того, чтобы лучше оснащать своих 4000 исследователей по всему миру в следующие несколько лет. Мы считаем, что L'Oréal не просто применяет ИИ в косметике — она дает ей свой собственный макияж. С такими данными, как у нее, и такой смелостью, как у ее помад, кто бы мог подумать, что ИИ может так хорошо подбирать цвета?

### Путь вперед: как создавать ценность с помощью ИИ

Мы считаем, что большинство компаний должны использовать сочетание всех трех моделей потребления ИИ. Вы будете использовать стороннее ПО с внедренным ИИ, и иногда будет вполне уместно использовать чужой ИИ-интерфейс для выполнения нужной вам задачи. Например, если вы агент по недвижимости и хотите получить быстрое описание кухни на основе полученных фотографий для нового объекта. Если у вас нет какого-то собственного уникального подхода к описаниям, это как раз тот случай, когда можно воспользоваться одной из известных моделей без особых опасений.

Но что если вы классифицируете тональность отзывов о покупке на основе тысяч мнений, собранных за три десятилетия продаж домов? Чтобы полностью раскрыть ценность ИИ и выделиться среди конкурентов, вам нужно использовать платформенный подход (как это делает L'Oréal), создавая ценность путем настройки ИИ под ваш бизнес. А другие модели потребления ИИ можно добавлять там, где это уместно.

Давайте подробнее рассмотрим создание ценности с помощью ИИ, начнем с LLM.

Напомним, что LLM (большие языковые модели) — это крупномасштабные глубокие нейросети, обученные на огромных объемах данных и далее адаптируемые под множество прикладных задач. Они могут быть как широкими и универсальными, так и более узкими и специализированными, но суть в том, что они предобучены с расчетом на то, что вы сможете дообучить их на своих собственных

данных, если хотите стать ИИ-создателем ценности. Это как новый сотрудник в вашей компании: он приходит с базовыми навыками и готовностью учиться. Чем больше он узнает о вашем бизнесе, тем больше накапливает институциональных знаний и опыта, и тем большую ценность он приносит (и тем более болезненной будет его потеря, если он уйдет к конкуренту).

С LLM все примерно так же. Вы используете свою ИИ-платформу, чтобы донастроить модель с учетом специфики вашего бизнеса, ваших собственных знаний и опыта — и она постепенно становится экспертом в вашей области и все ценнее для вас со временем. Вы ведь не захотите, чтобы ваш обученный на внутренней информации продавец ушел к конкурентам — ИИ-создатели ценности относятся к своим данным точно так же!

А поскольку ИИ-создатели ценности контролируют платформу, процессы и данные, они со временем накапливают все больше ценности. Мы уже видели, к чему приводит потеря контроля в случае потребительских ИИ-сервисов: это и плохие данные, ведущие к неверным результатам, и конфабуляции (галлюцинации модели), и даже проблемы с авторскими правами из-за неосознанного использования охраняемого контента (именно на этом основаны текущие судебные иски). В некоторых случаях даже наблюдалась утечка конфиденциальных данных в публичные пространства.

Вот почему, говоря о корпоративном ИИ, крайне важно понимать, как именно была построена ваша LLM, какие данные использовались для обучения и по какому «рецепту» она была собрана. А также — почему нужно жестко контролировать все данные, особенно чувствительные. Сильное управление ИИ (AI governance) абсолютно критично.

### [Смотри, прежде чем прыгать](#)

Да, сейчас самое время включаться в ИИ, но сначала посмотри под ноги. Убедись, что ты инвестируешь в разумный, безопасный и устойчивый подход, при котором основными бенефициарами становятся твой бизнес и клиенты. Мы считаем, что этот путь начинается с ИИ-создателя ценности, использующего надежную платформу.

### [Планирование ИИ-будущего: в нем множество GenAI-моделей](#)

Сейчас, по нашему мнению, существует миф об ИИ — или, по крайней мере, базовое недопонимание. Для широкой публики кажется, что GenAI появился буквально из ниоткуда. Многие думают, что существует лишь горстка ориентированных на потребителя ИИ-сервисов и что победит одна-единственная модель (в духе «одна модель, чтобы править всеми», если вспомнить Толкина).

Мы не думаем, что так будет. Будущее ИИ — это не одна модель. Это множество моделей (иногда это называют многомодельным подходом), и при этом они еще и мультимодальные (то есть работают с изображениями, текстами, видео, звуком и так далее). Твой бизнес будет использовать несколько дообученных моделей, чтобы достигать наилучших результатов в конкретных сценариях применения. Некоторые модели будут готовыми «с полки», некоторые — настроены на твоих данных, некоторые будут использоваться для оценки результатов других моделей (они называются judge models), некоторые будут использоваться как есть — для обеспечения безопасности, другие — для задач с необходимостью сложных рассуждений, и еще часть — для создания агентов. Именно поэтому платформенный подход так важен — и поэтому мы в главе 1 познакомили тебя с сообществом Hugging Face.

И как мы уже намекали (а на самом деле прямо сказали, просто вежливо) — делай ставку на сообщество, потому что будущее ИИ — это не столько про закрытые модели, сколько про развитие через открытую науку и открытый исходный код. Закрытые модели, конечно, будут играть свою роль, но большая часть того, что будет происходить в будущем, не должна — и не будет — происходить за закрытыми дверями. Все должно происходить на виду, с полной прозрачностью и подотчетностью в рамках open source.

Повторим: сейчас в open source-сообществе вокруг GenAI и агентов царит невероятная энергия. Есть распределенные проекты, университетские инициативы, корпоративные усилия — все это движет инновации вперед и создает LLM, которые ты можешь донастраивать и использовать под свои задачи.

Многие говорят: «ИИ от Big Tech — это проблема». Мы с этим не согласны (и не потому, что работаем в крупной технологической компании). Мы бы предпочли, чтобы вы шире посмотрели на ситуацию и сказали: «Закрытый и проприетарный ИИ — потенциально серьезная проблема». С этим мы согласны. Почему мы на этом акцентируем внимание? Потому что есть поставщики как крупные, так и

небольшие (мы не будем называть их поименно... мы не ищем конфликта), которые используют закрытые и проприетарные подходы. А есть компании, крупные (например, IBM и Meta) и небольшие (например, Mistral AI и DeepSeek и другие), которые придерживаются открытости.

В долгосрочной перспективе, в интересах всего общества, мы не хотим, чтобы выиграла лишь один или несколько игроков — чтобы несколько компаний определяли, что такое ИИ, и диктовали, как его использовать. Судя по тому, что мы видим, этого не произойдет — и это отличная новость для вас, вашего бизнеса и общества в целом.

### *Время демистифицировать ИИ и начать применять его*

Как когда-то говорили, что данные — это «новая нефть», теперь ИИ называют «новым электричеством» мира. Помимо того, что GenAI делает современный ИИ повсеместным и все более доступным (благодаря prompt-интерфейсу), ИИ может (и будет, если все сделать правильно) усиливать и трансформировать способы ведения бизнеса по всему миру. Сегодня ИИ может использоваться для создания сверхточных прогнозов, автоматизации бизнес-процессов и принятия решений. Влияние огромно: от бесшовных клиентских взаимодействий до интеллектуальных продуктов и более эффективных сервисов. Это приведет к экономическому эффекту для компаний, стран и обществ в целом.

Очевидно, что организации, которые наладят масштабные эксперименты с ИИ, станут победителями в следующем десятилетии рыночных возможностей. Чтобы разобраться и помочь демистифицировать ИИ, нужно рассматривать две ключевые составляющие этой категории: компоненты и процесс. Иными словами, нужно понять, что стоит за ИИ и как его можно внедрять.

### **Компоненты**

Точно так же, как развитие использования электричества было обусловлено базовыми компонентами вроде резисторов, конденсаторов, диодов и так далее, развитие ИИ движется вперед благодаря современному программному компонентному подходу, который включает в себя элементы, описанные в этом разделе.

### **Единая, современная data fabric-архитектура с подходом «данные как продукт»**

Вы уже не раз слышали от нас в этой книге: вашему ИИ нужна ИА (информационная архитектура, IA). Почему? Потому что ИИ питается данными, и, следовательно, ваши данные должны быть подготовлены для ИИ. (Именно поэтому это стоит первым в списке.) Это выходит за рамки принципа «мусор на входе — мусор на выходе» (GIGO), хотя с ИИ эта проблема становится еще более острой, ведь все, что он делает, — это находит числовые закономерности, о которых мы упоминали в главе 1. Будет проблемой, если вы считаете, что в интернете все — правда, нет фейковых новостей, ненависти, оскорблений или нецензурной лексики. Иными словами, проблема реальна.

Правильно реализованная data fabric охватывает все данные предприятия с обеспечением управляемого поиска и связанности. Она устраняет сложность подключения к данным и понимания деталей лежащих в основе технологий с помощью data intelligence. Мы часто повторяем: «Облако — это возможность, а не пункт назначения!» (Гибридное облако — это подход, с которым сегодня уже согласилось большинство компаний.) Точно так же вы используете data fabric, чтобы применить аналогичное мышление к своей ИА: «данные не находятся только в одном месте» — и это мышление дает преимущества повсеместно.

Data fabric выступает как логическое представление всех активов данных в любом облаке (публичном, приватном или локальном). Она автоматически организует и маркирует данные по всему предприятию (и за его пределами, если нужно), независимо от их местоположения. Она обеспечивает подход, при котором операции доставляются к данным, а не наоборот, что оптимизирует циклы вычислений. Говоря проще, это означает, что нужные действия выполняются там, где находятся данные, а не все данные стягиваются в одно место для обработки. В мире больших данных вы легко поймете, почему второй подход не масштабируется.

И, возможно, самое важное — она предоставляет сотрудникам компании управляемый и бесшовный доступ ко всем доступным данным через виртуализацию, от фаервола до периферии. Когда вы думаете о data fabric, думайте о самообслуживании, удобстве доступа и защите данных. В конечном счете, data fabric трансформирует использование данных в процесс «сшивания» информации по всему вашему бизнесу — и за его пределами, если это уместно.

## Данные как продукт

Хотя это выходит за рамки данной книги, мы не могли не упомянуть концепцию «данных как продукта», поскольку она тесно связана с архитектурой data fabric. (Возможно, вы уже слышали о data mesh, так вот — это, по сути, начальная форма идеи «данные как продукт».) Data mesh — это подход, при котором на данные начинают смотреть как на продукт. Эта архитектура является не столько технологической, сколько культурной: она смещает ответственность за достоверность данных с ИТ-команд на бизнес-команды, которые владеют и курируют данные.

«Данные как продукт» означают, что данные рассматриваются как API: каждая бизнес-единица (иногда называемая доменом) несет ответственность за то, чтобы данные, стоящие за их API, были качественными и доступны другим бизнес-единицам. Когда вы думаете о данных как о продукте, вы сталкиваетесь с еще одним важным принципом эффективной информационной архитектуры, который помогает ИИ: владение доменом. Это означает, что бизнес-единицы берут на себя ответственность за свои данные.

Еще один компонент подхода «данные как продукт» — это федеративное управление данными. Речь идет о согласованном управлении данными по всем источникам и об использовании автоматизации (в том числе с помощью ИИ) для поддержки этой задачи. Именно поэтому мы считаем, что подход «данные как продукт» отлично сочетается с архитектурой data fabric. Многие компании пытаются создать этот компонент самостоятельно вместо использования готового решения, и в результате теряют время, деньги и срывают сроки проектов.

Когда вы думаете о данных как о продукте, думайте о курировании, управлении (при поддержке data fabric) и отслеживании происхождения данных.

Эффективная стратегия информационной архитектуры включает больше элементов, чем перечислено выше, но именно эти — ключевые рычаги, с которых нужно начинать. После этого задачи по сбору, организации, управлению, внедрению данных в существующие ИИ- и не-ИИ-процессы и другие аспекты встанут на свои места.

## Среда разработки и движок

Бизнесу необходимо пространство, где он может создавать, обучать, масштабировать и запускать свои модели ИИ. В идеале все эти компоненты должны быть интегрированы с вашими стратегическими решениями по хранению данных (например, через облака данных) и с рамками управления — и все это должно быть связано через общее пространство в рамках экосистемы. Такой подход помогает организациям выстраивать общее понимание целей, языка и процессов проектирования — от ввода до вывода. Когда у вас есть и data fabric, и подход «данные как продукт», ваша стратегия работы с данными начнет ощущаться как магия. И хотя мы разоблачили миф о «волшебстве ИИ», наличие четкого плана и инерции за спиной действительно вдохновляет.

## Модальность человеческих особенностей

Один из способов «оживить» ИИ-модели — это соединить их с человеческими характеристиками, такими как голос, язык, зрение и рассуждение. GenAI, особенно в виде агентов, часто упоминается в контексте бесшовного клиентского опыта, где обычно обсуждаются чат-боты. Однако термин «чат-бот» часто ассоциируется исключительно с текстовой перепиской, хотя голосовой интерфейс — это такая же модальность. Многие из нас сталкивались с IVR (автоответчиками), которые звучат неестественно, но с помощью ИИ можно придать голосу реальное человеческое звучание и выразительность. Например, вы можете загрузить документ в Google NotebookLM и попросить его сгенерировать подкаст — результат впечатляет. ИИ позволяет модернизировать IVR, добавляя выразительные голоса с элементами живой речи: интонацией, эмоциями, акцентами и междометиями.

Хотя в книге рассматривается агентный ИИ, мы не фокусируемся на влиянии агентов на модальность взаимодействия, особенно с точки зрения UX. Будущие UX-дизайнеры должны будут учитывать два типа пользователей: людей и агентов. Опыт агента (AX) — это уже не только API, но и настольные взаимодействия.

Эта способность важна, потому что мы, люди, передаем эмоции в произносимых словах. Мы можем звучать сочувственно, когда извиняемся, неуверенно — когда не знаем ответа, и радостно — когда делимся хорошими новостями. Именно передача эмоций делает наш голос человеческим. ИИ может использовать это, чтобы снизить уровень разочарования клиентов в телефонных взаимодействиях.

Но главный момент вот в чем (и почему переобучение становится такой важной темой): кастомизированные брендовые голоса (включая ваш собственный) сегодня можно создать за считанные минуты — без технической подготовки. Выразительный голос создает у клиента ощущение общения с живым человеком, а не с роботом, но ваша компания при этом получает выгоду от «смещения влево»: перевод простых запросов с живых агентов (стоимость — около \$5 за взаимодействие) на ИИ-ассистентов (стоимость — около \$0.25 за взаимодействие). Это отличный пример типичных задач, которые мы игнорируем ежедневно, но которые легко улучшить с помощью технологий. Если вы отвечаете за поддержку с IVR и не знаете, как просто создать взаимодействие с естественным голосом — вы продолжаете мириться с лабиринтом «Нажмите 1, чтобы...», где вместо приза в конце ваши клиенты в отчаянии кричат «Свяжите меня с оператором!».

Это приводит нас к мультимодальному ИИ, в котором человеческие черты становятся все более заметными. Например, Google Gemini, Apple FERRET, Meta Llama, DeepSeek Janus-Pro, IBM Granite и разные модели OpenAI позволяют включать изображения в подсказки и объяснять, что на них изображено. Представьте, что служба доставки прислала вам фото вашей посылки у двери, а ИИ написал, например: «Один из углов коробки поврежден». Также может быть автоматически подготовлена форма для подачи претензии, если содержимое окажется испорченным. Если все в порядке — ничего делать не нужно. Если нет — возврат будет максимально бесшовным. Это и есть работа агентного ИИ.

Представьте себя на месте клиента в этом сценарии. Конечно, никто не хочет, чтобы что-то пошло не так (будь то поврежденный товар, неправильная доставка или сброс пароля), но важнее другое: если уж проблема возникла, клиент не должен сталкиваться с препятствиями — многократными переводами, повторной идентификацией и другими моментами, из-за которых мы каждую неделю удивляемся: «Что за черт?». Исследования показывают, что по-настоящему качественный клиентский опыт не всегда в том, чтобы все шло идеально. Бизнес может ошибаться — и клиенты это готовы терпеть (разумеется, не в критических отраслях вроде кардиохирургии). Главное — обеспечить отсутствие трения. Это обеспечивает лояльность, терпение и доверие, даже когда все идет не по плану.

### Управление и эксплуатация ИИ

Этот компонент позволяет с уверенностью внедрять ИИ в любые приложения и бизнес-процессы. Но чтобы это сделать правильно, необходимо понимать, как модель была построена, какие данные использовались, как повысить ее эффективность, что изменилось с течением времени, а также учитывать такие параметры, как дрейф, смещение (bias) и дисперсия (variance). Именно здесь размещаются модели для их эксплуатации и полного управления жизненным циклом ИИ. Кроме того, этот элемент обеспечивает доказуемость и интерпретируемость решений, принимаемых ИИ.

Представьте себе следующее: если бы мы попытались зафиксировать объем данных, генерируемых в мире каждую минуту, эта цифра устарела бы уже в момент сохранения черновика. Каждый раз, когда мы бы обновляли эту главу, она снова становилась бы неактуальной. С ИИ-моделями — та же история: они начинают «дрейфовать» с момента вывода в продакшн. Если данные, на которых обучалась модель, больше не «рифмуются» с актуальными данными, модель начнет уходить от первоначального назначения (например, выбор рыночной возможности) и/или может начать генерировать некорректные результаты, включая усиление предвзятости. Это и есть проблема дрейфа, которую необходимо отслеживать и управлять ею.

### Агенты и ассистенты для массового использования

По мере того как ИИ интегрируется в «нервную систему» бизнеса, классифицируется и привязывается к рабочим процессам (это и есть концепция +AI), стоит помнить, что агенты и ассистенты — ключевые инструменты, с помощью которых ИИ начинает приносить реальную пользу. Именно они позволяют демократизировать ИИ в масштабах всей компании.

Да, крайне важно иметь ИИ-платформу, которая позволяет собирать, организовывать и хранить данные, строить модели GenAI и управлять ими. Но именно агенты и ассистенты являются «шасси» — они дают возможность использовать мощность этих моделей на практике и масштабировать ценность ИИ для всего предприятия. Например:

- Команды разработчиков могут использовать Microsoft Copilot или IBM watsonx Code Assistant, чтобы ускорить процесс разработки.

- Вы можете проектировать бесшовный клиентский опыт с помощью watsonx Assistant или Kore.ai.
- А можно даже оркестрировать бизнес-процессы с помощью Aisera или watsonx Orchestrate и его библиотеки ИИ-агентов.

Эти примеры демонстрируют реальное повышение продуктивности сотрудников с помощью ИИ. И мы считаем это критически важной частью любой успешной ИИ-стратегии, потому что она отвечает на два главных вопроса: **кто будет использовать ИИ и как он им поможет?** В зависимости от вашей роли в организации вы должны либо знать ответы на эти вопросы, либо хотя бы уметь их правильно сформулировать.

### Процесс: ингредиенты для торта без рецепта не дают торт

Имея на руках все перечисленные ранее компоненты, организации смогут раскрыть ценность, скрытую в их данных. Однако для того чтобы по-настоящему использовать потенциал ИИ, необходимо понимать, как его внедрять и применять. Ниже приведены базовые шаги, которые помогут запустить ИИ в бизнесе (детали — в следующих главах книги).

**Первый шаг** — определить правильные бизнес-задачи для применения ИИ. Потенциальных направлений много: обслуживание клиентов, повышение производительности сотрудников, выявление производственных дефектов, оптимизация затрат в цепочках поставок и так далее. Все, что можно четко описать, можно запрограммировать, а значит, ИИ может это улучшить. Как уже обсуждалось в первой главе (и будет более развернуто в четвертой), возможности практически безграничны, но крайне важно фокусироваться на задачах и результатах бизнеса, а не на проектах ради самой науки о данных. В эпоху ажиотажа вокруг Hadoop многие компании вкладывали большие бюджеты и время в проекты, которые не давали реальной бизнес-ценности или не были пригодны к использованию. Именно поэтому GenAI так сильно отличается: он позволяет быстрее создавать кейсы и делает их доступными широкой аудитории. Но при этом важно правильно выбрать приоритеты.

**Второй шаг** — подготовить организацию к использованию ИИ. Необходима будет серьезная прокачка компетенций — от создания команд специалистов по data science до формирования более широкого взгляда на то, как именно GenAI может трансформировать бизнес. Это поможет избежать ситуации, когда вы каждый день проходите мимо проблем, которые давно могли бы быть решены с помощью технологий.

Придется активно прокачивать знания сотрудников по GenAI, большим языковым моделям и агентам. Речь не идет о том, чтобы устраивать маркетологу экзамен по регрессии Lasso или по метрике AUC (area under the ROC curve). Нужно, чтобы у всех была базовая осведомленность о возможностях и рисках ИИ — это критично для его эффективного применения. Особенно важно: у вас должен быть план по развитию всех сотрудников, чтобы выгоды от ИИ распределялись по всей компании. Этому посвящена целая глава книги — шестая.

Почему это так важно? Потому что многие рутинные и ручные задачи будут автоматизированы (смещены влево), и это изменит роли сотрудников. Редко когда роль может быть полностью заменена ИИ, и столь же редко она не может быть дополнена ИИ. Любая технология бесполезна без людей, которые смогут ее применить, поэтому нужно создавать ядро экспертов, которые будут вдохновлять и обучать других, и параллельно — поддерживать постоянное развитие навыков у остальных. Ведь если обычный технологический прогресс сопоставим с возрастом собаки (1 год ИТ равен 7 годам), то GenAI и агенты развиваются с мышинной скоростью (1 год — как 30 лет), и без четкого плана просто не угнаться.

**Третий шаг** — выбрать технологии и партнеров. Хотя маловероятно, что генеральный директор лично будет выбирать стек GenAI для компании, в этом пункте важен культурный аспект. Организации стоит пробовать разные технологии, сравнивать их и учиться в процессе. Один практичный совет, который избавит от боли: не попадайтесь в ловушку мышления, что облако — это один вендор, одна платформа. Опыт последних лет показал, насколько ошибочной была эта идея. Это не значит, что вам нужно иметь сотни поставщиков ИИ (их и правда становится все больше), но не существует одной модели ИИ, которая решит все. Компаниям нужно выбрать несколько надежных партнеров, обладающих как технологиями, так и компетенциями для реализации ИИ. Слово «надежных» здесь ключевое. В мире технологий вы, скорее всего, уже сталкивались и с «положительными героями», и с

«токсичными участниками». Мы считаем, что именно доверие станет главным лицензирующим фактором, и предлагаем вам самостоятельно задуматься, кому вы доверяете.

Успех чаще всего приходит через партнерства — будь то деловые или личные. Вспомните: у Бэтмена был Робин, у Шерлока — Ватсон, у Берта — Эрни. Даже у Снуки была Ситуация (и если вы не знаете, кто это, мы даже рады — меньше лишнего багажа).<sup>1</sup>

### Принимайте провалы, но делайте это безопасно

Вы знали, что более 80% традиционных AI-проектов никогда не доходили до продакшена? GenAI должен улучшить эти показатели благодаря простоте запуска, но вы все равно столкнетесь с трудностями и неудачами — от неправильных ответов до юридических ограничений. Возможно, вы попробуете 40 проектов, и 30 из них не сработают. Но если 10 окажутся успешными и были выбраны правильно, они перекроют убытки от остальных — об этом как раз подробно рассказывается в четвертой главе.

Многие любят повторять: «Проваливай быстро и иди вперед». Но в контексте GenAI и особенно агентов этого недостаточно. Подумайте: сказали бы вы то же самое своему ребенку, обучающемуся в университете, за обучение которого вы платите? Вряд ли. Лучше использовать подход: «Проваливай быстро, иди вперед и делай это безопасно». Поэтому мы считаем, что регулировать нужно не сам ИИ, а конкретные сценарии его применения. Согласитесь, что система, влияющая на уголовные приговоры, должна регулироваться намного жестче, чем ИИ, предлагающий вам очередной сериал на вечер. Один промах — и вы посмотрите «Домохозяек из Нью-Джерси», другой — и человек окажется за решеткой.

Это еще и вопрос культуры. В компании должно быть принято безопасно ошибаться, делать выводы и двигаться дальше. И для тех, кто руководит AI-проектами (а значит, и принимает возможность неудач), мы предлагаем зафиксировать в памяти одну мысль — это смесь высказываний Майкла Хайетта и Форреста Гампа: «В твои лучшие дни ты, возможно, не так умен, как тебе кажется. Но в худшие — точно не так глуп, как ты думаешь».

### Будущее ИИ

Со всеми достижениями последних лет амбиции 1950-х годов вновь становятся актуальными. Современные модели еще не являются истинным искусственным общим интеллектом (AGI), хотя модели с элементами рассуждения приближают нас к этому), но некоторые из них уже могут пройти тест Тьюринга (изначально называвшийся «игрой в подражание»), который проверяет способность машины демонстрировать интеллектуальное поведение, неотличимое от поведения человека. Что это значит для всех нас?

Некоторые, сталкиваясь с генеративным ИИ и агентами, считают, что мы стоим на пороге светлого утопического будущего, в то время как другие видят в этом предвестие мрачной антиутопии. Мы придерживаемся умеренной, но скорее позитивной позиции: технология не обязательно должна уничтожить мир, чтобы изменить его. Как мы уже говорили в первой главе, не должно быть сюрпризом, что технологические инновации могут и помогать, и вредить нам (хороший пример — социальные сети). Мы хотим, чтобы вы понимали: и оптимизм, и тревожность — это обоснованные реакции, ведь общество сомневалось в каждой крупной технологической вехе с начала промышленной революции (и во многих случаях ошибалось в своих прогнозах).

---

<sup>1</sup> ChatGPT поясняет:

Снуки (Snooki) и Ситуация (The Situation) — это прозвища участников американского реалити-шоу "Jersey Shore", выходявшего на MTV с 2009 года. Шоу рассказывало о жизни группы молодых людей, проводящих лето в Нью-Джерси (а потом и в других местах), и быстро стало культурным феноменом — во многом из-за эксцентричного поведения, вечеринок, конфликтов и гипертрофированной "гедонистической" эстетики.

- Снуки (настоящее имя — Николь Полици) — одна из самых ярких участниц шоу. Невысокая, громкая, взрывная, с очень узнаваемым образом. Ее поведение и манера говорить сделали ее символом всего проекта.
- Ситуация (настоящее имя — Майкл Соррентино, Mike "The Situation" Sorrentino) — другой известный участник. Его прозвище происходит от того, как он хвастался своим прессом, называя его «ситуацией». Он был известен своим нарциссизмом, частыми ссорами и "мачо"-повадками.

Упоминание этих персонажей в тексте — это ироничный пример того, что даже у самых хаотичных или карикатурных персонажей есть партнеры. Сравнение подчеркивает идею, что все успешные "герои", даже в поп-культуре, чего-то добиваются через сотрудничество — будь то детектив и помощник или звезда реалити и ее шоу-партнер.



ИИ будет влиять не только на наш цифровой мир. Он также затронет наш физический мир; и при правильном применении представьте, как ИИ может ускорить открытия и инновации. Это касается не только макияжа; подумайте, как ИИ может помочь в создании новых материалов для медицины, энергетики, климата и других насущных вызовов, стоящих перед человечеством — это те же самые задачи, что и в макияже, просто описанные на другом «языке». А по мере развития квантовых вычислений мы увидим синергию этих инноваций, которые можно будет использовать для решения этих и других задач.

Наконец, что насчет совершенно нового типа вычислений на базе генеративного ИИ — об этом мы расскажем в девятой главе. В конечном счете наш успех — и успех всего человечества — зависит от того, как мы и остальной мир подойдем к ИИ.

### Давайте перейдем к делу

Мы уже затронули множество тем (на высоком уровне), но, по сути, мы говорили вам, что для успеха в ИИ нужно проделать массу «не технарской» работы. Возможно, это звучит обескураживающе. Мы не хотели, чтобы у вас возникло такое чувство, но небольшое внутреннее беспокойство необходимо, чтобы вы двигались быстрее — с намерением, чтобы не упустить момент. Цель оставшейся части книги — устранить барьеры на вашем пути, а не создавать их.

Не заблуждайтесь: если вы ощущаете срочность и страх упустить момент, это нормально. Можем вас заверить, что почти каждая компания находится в такой же ситуации, и множество людей сейчас испытывают те же эмоции, что и вы. И поверьте, мы слышали массу «рыбацких историй» от отдельных лиц и компаний, рассказывающих об их ИИ или о том, как их продукты построены с помощью ИИ, — и, как в большинстве рыбацких историй, многие из них преувеличены или неправдивы. Нам хочется сказать таким людям: «Не стоит рассказывать сказки тем, кто знает реальный размер рыбы», — но мы просто улыбаемся и продолжаем свой день. Тем не менее, после прочтения этой книги вы будете в такой же позиции, и мы оставим вам решение — просто улыбнуться или нет.

Обещаем вам (и вашему бизнесу) следующее: если вы проявите сдержанность и не поставите галочку «мы внедриli ИИ в бизнес» с помощью быстрых и легких решений (или под давлением); если вместо этого вы будете продуманы, последовательны и стратегичны в использовании платформы, которая учитывает все необходимые компоненты (ИИ, управление данными, интеграция данных и контроль); и что еще важнее — если вы установите свой «навигационный маршрут» на цель «Создатель ценности с помощью ИИ», — тогда вы будете в отличной позиции для долгосрочного успеха. Более того, как и многие до вас, ваша компания не будет начинать все сначала каждый раз, когда направление ветра в ИИ меняется.

Лично мы очень воодушевлены этой новой главой в технологии. Мы — все вместе — собираемся использовать генеративный ИИ и агентов, чтобы преобразить не только цифровой, но и физический мир. Мы собираемся использовать его, чтобы справиться с самыми сложными социальными, медицинскими, экологическими и другими проблемами. Мы сделаем это с помощью науки, но также и через расширение возможностей бизнеса — таких, как ваша компания и наша — делать больше, быстрее и ответственнее. Чем бы ни занималась ваша компания, ИИ станет мощным новым инструментом, который поможет вам делать это лучше.

Мы абсолютно уверены в следующем: именно Создатели Ценности с помощью ИИ окажут наибольшее влияние. Они возьмут удивительную фундаментальную технологию генеративного ИИ и используют ее для создания совершенно новых решений и рабочих процессов. Вот почему наша цель — сделать ИИ доступным для всех и передать его вам.