

Глава 4. О практическом применении

Это продолжение перевода книги Томас и др. Создание бизнес-ценности с генеративным ИИ. Существует множество исследований, подтверждающих колоссальный потенциал повышения продуктивности благодаря генеративному ИИ, особенно агентам. Если вспомнить формулы из предыдущей главы, вы уже можете почувствовать, какие перспективы открываются. Но насколько они масштабны? Попытка это оценить — все равно что выбирать между проходкой за кулисы на концерт Metallica или ролью в следующем фильме про Дэдпула: в любом случае будет эпично, но невозможно заранее понять, от чего сердце забьется чаще.

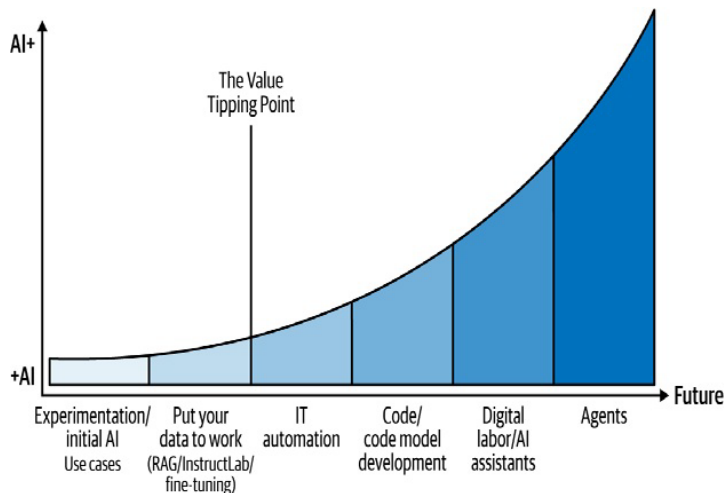


Рис. 4.1. Компании, переходящие от +AI к AI+, обычно следуют схожему пути создания ценности — кривой создания ценности с помощью ИИ

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

Допустим, что генеративный ИИ может высвободить более 10 трлн долларов дополнительной продуктивности в ближайшие годы и до 20 трлн к концу десятилетия (это наш консенсус на основе агрегирования всех встречавшихся нам аналитических отчетов). Какую часть из этого сможет получить ваша компания? И это только начало: мы убеждены (как, впрочем, и большинство пишущих об ИИ), что впереди появятся совершенно новые, ранее немыслимые бизнес-модели и продукты.

Вспомните: идея ночевать в доме незнакомцев во время командировки казалась абсурдной — до тех пор, пока в Сан-Франциско не возник Airbnb. Так почему бы не использовать уже доступные преимущества ИИ для того, чтобы протестировать и улучшить сценарии его применения, модели управления, обучение, работу с данными — и все это в рамках нового взгляда на будущее?

Как лидеры бизнеса, мы постоянно сталкиваемся с необходимостью принимать важные решения о том, как обеспечить отдачу от инвестиций. Сейчас мы стоим на пороге ключевого момента — пересечения путей гибридных облаков и ИИ. Поскольку эти технологии выходят на зрелый уровень, нам необходимо внедрять их, чтобы создать конкурентное преимущество и запустить волну инноваций.

Этот момент сравним с эпохой Netscape: он дает каждому из нас возможность переосмыслить технологический ландшафт — не только чтобы максимально использовать потенциал ИИ, но и чтобы быть готовыми к тем технологиям, что придут вслед за ним. Суметь собрать воедино правильные элементы — вот что определит, будете ли вы лидером в своей отрасли или останетесь позади.

Бизнес-ИИ — это история про создание ценности, а ценность рождается из правильных сценариев применения. Сегодняшние решения окажут глубокое влияние на завтрашний день. В этой главе мы рассмотрим больше сценариев и расширим рамки классификации, о которой говорили в первой главе. Конечно, про сценарии в рамках одной только отрасли можно написать целую книгу, но наша цель здесь — укрепить вашу насмотренность и дополнить модель классификации сценариев. Мы укажем, с чего стоит начинать, покажем больше отраслевых примеров и прокомментируем горизонтальность (в смысле межотраслевое распространение) сценариев использования генеративного ИИ и агентов — все для того, чтобы вы смогли раскрыть внутренний потенциал вашего бизнеса.

Лучший способ найти примеры использования генеративного ИИ в вашей отрасли — это просто сделать обычный (или с помощью GenAI) веб-поиск. Публикаций, статей и аналитики по отраслевым кейсам — буквально неисчерпаемое количество. Консалтинговые компании вроде IBM, Deloitte, McKinsey и другие регулярно публикуют материалы на эту тему, как и технологические компании и практики. Мы говорим вам это не потому, что сами не можем все охватить. Если честно, нам даже иногда попадаются продавцы, которые спрашивают у нас про сценарии для своей отрасли. В таких случаях мы делаем скидку на то, что они, возможно, не знают, что можно спросить у Google или GenAI. Потому что если бы знали — сказали бы прямо: лень. Звучит жестко, но настолько много информации уже доступно. Вы удивитесь, насколько быстро можно вникнуть в ИИ-кейсы любой отрасли с минимальными усилиями.

Напомним: формула успеха в ИИ из главы 3 — это сложение четырех компонентов: модели + данные + управление + сценарии. Если LLM, данные и управление — это «лесá», на которых все держится, то сценарии применения — это сам «дворец», ради которого все строится — именно они создают ценность.

Кривая создания ценности в сценариях применения

До этого момента мы уже дали вам несколько инструментов, которые помогут определять проверенные и эффективные кейсы. Как у электрика есть набор инструментов (стрипперы, бокорезы, отвертки с изоляцией, тестеры и т.д.), так и мы даем вам инструменты для вашего «ИИ-пояса», чтобы вы могли с уверенностью вести свою организацию к результатам, которые «ударят током» — но в хорошем смысле.

Давайте на минуту остановимся и вспомним, что уже есть в вашем наборе. В первой главе мы предложили кривые зрелости для оценки кейсов, схему классификации по бюджету (тратить, чтобы зарабатывать, или тратить, чтобы экономить), а также дополнительную классификацию по функционалу ИИ (автоматизировать, оптимизировать или предсказывать). А сейчас дадим вам еще один инструмент — он поможет выбрать ИИ-сценарии с наибольшим потенциалом для вашей компании.

Посмотрите на рис. 4.1 и представьте себя в роли знаменитого британского лыжного прыгуна и героя-аутсайдера Эдди «Орла» Эдвардса — только этот трамплин запускает вас в ИИ-будущее. Начальный участок трамплина — это фаза осторожного старта (Эксперименты). На этом этапе вы набираете скорость — в прыжках с трамплина это называется разгонный спуск — и здесь особенно важно соблюсти баланс. Не перегибайте: помните принцип fail fast, fail forward, fail safe (провались быстро, двигайся вперед, делай это безопасно). И хотя многие компании, которые в итоге «взлетели», начинали именно с экспериментов, слишком многие все еще там застряли (и это одна из причин, по которой мы вообще написали эту книгу).

Далее, по мере накопления опыта, вы выходите в «аэродинамическую стойку» и начинаете по-настоящему использовать данные, применяя техники вроде настройки моделей (InstructLab, LoRA и др.), мультимодального промптинга или паттерна RAG (retrieval augmented generation, генерации дополненной извлечением) — мы подробнее обсудим все это в 8-й главе. И вот вы достигаете ключевого момента — точки перегиба ценности (Value Tipping Point)!

В отличие от прыжков с трамплина, здесь есть возможность повернуть назад — но зачем? На этом этапе у тебя уже накоплена инерция. Подумай сам: за плечами есть успешные кейсы, ты понимаешь возможности GenAI и агентных систем, у тебя есть четкое представление о критериях оценки ИИ-платформы: выбор, эффективность и прозрачность. И (надеемся) благодаря этой книге ты уже подключился к открытому инновационному ландшафту (выбор), у тебя есть LLM-модели, которые работают на уровне требований бизнеса (эффективность) и безопасны для внедрения (доверие). Проще говоря — ты уже не можешь позволить себе отступить. Если сделаешь это, ты лишишь свой бизнес возможности реализовать весь потенциал.

Ведь именно справа от точки перегиба на кривой создания ИИ-ценности начинается настоящее ускорение: автоматизация и цифровой труд становятся катализаторами для других улучшений — например, в разработке кода, в снижении затрат с помощью ИИ-ассистентов.

На самом конце этой «трамплинной» траектории ты действительно взмываешь в воздух — и, как олимпийский прыгун с трамплина, ты теперь в идеальном положении: максимальный подъем,

минимальное сопротивление. Ты стал агентным. Все, над чем ты работал, стало частью операционной модели компании — у тебя есть агенты, ассистенты, ты используешь свой самый ценный актив (данные), ты автоматизируешь рутинные задачи во всех функциях — от разработки и маркетинга до копирайтинга. Правый край кривой — это и есть точка назначения: AI+ достигнуто.

И даже если сначала ты не выглядишь как профессионал мирового класса, как и «Орел» на старте своей карьеры, — после прыжка все зависит от того, насколько высоко и далеко ты полетишь. А главное — благодаря тому, что ты прочитал эту книгу, мы уверены: твоё «приземление» (если оно вообще нужно — а мы считаем, что нет) будет куда удачнее, чем у Эдди в его первые попытки.

Горизонтальные сценарии дают максимальный вертикальный эффект

На рис. 4.1 ты заметишь, что сценарии (и техники) не являются отраслевыми (вертикальными), а наоборот — охватывают сразу несколько отраслей, то есть являются горизонтальными. Мы будем повторять это снова и снова, потому что считаем: при выборе ИИ-сценариев сначала нужно мыслить горизонтально, а уже потом — вертикально.

Давай теперь, без всяких лыжных метафор, просто рассмотрим этапы кривой создания ИИ-ценности.

Эксперименты

Для большинства первые проекты в области ИИ начинались с data science и машинного обучения. Но поскольку эти подходы требовали глубокой экспертизы и зачастую — значительных усилий по разметке данных, получаемая ценность была ограниченной и не масштабировалась на всю компанию. Сценарии применения были разными: часть — горизонтальными (исследование данных, клиентская поддержка, персонализация в маркетинге), часть — отраслевыми (мошенничество в платежах, аналитика в ритейле и т. д.). Этот этап развития ИИ так и не смог масштабироваться.

С появлением генеративного ИИ начался новый цикл экспериментов. Вдруг стало возможно достичь экспоненциальных результатов, поскольку теперь не нужно столько усилий для подготовки и разметки данных. Сегодня многие убеждаются, что правильные данные в сочетании с подходящими моделями способны дать впечатляющие результаты. Проще говоря, экспериментировать с GenAI оказалось значительно проще, чем с традиционным ИИ, и именно поэтому многие, кто начал, быстро разобрались и начали применять GenAI на практике.

Используй свои данные с умом

Парадоксально, но самый ценный актив для GenAI во всех компаниях одинаков — это собственные (проприетарные) данные. Когда бизнесы стремятся извлечь выгоду из ИИ, вывод один и тот же: если мы сможем раскрыть потенциал своих внутренних данных, у нас появляется реальный шанс на экспоненциальный эффект (подробнее об этом мы поговорим в главе 8).

Шпаргалка по LLM

Если ты хочешь получать более достоверные результаты от LLM, ее нужно дополнить твоими знаниями. Представь, что ты работаешь в HR и используешь LLM как движок для чат-бота по кадровой политике, а также для рассылки писем. У тебя есть внутреннее корпоративное руководство, в котором подробно описаны правила семейного отпуска, больничных, программ благотворительности, политики удаленной работы и так далее. У стандартной LLM, обученной на публичных данных из интернета, просто не будет доступа к этим сведениям — а публиковать внутренний гайд в открытом доступе ты, скорее всего, не хочешь.

В корпоративной среде дополнение модели своими данными дает массу преимуществ и снижает риск «галлюцинаций» (об этом подробно в следующей главе; но в двух словах: LLM может просто придумать информацию — например, сказать, что отпуск по уходу за ребенком составляет 2 недели, когда на самом деле — 26).

Использование своих данных — это процесс встраивания их в работу GenAI (в безопасной, контролируемой и объяснимой форме), чтобы улучшить результат. Один из способов — это RAG (retrieval augmented generation). В этой схеме модель извлекает информацию из доверенного источника (например, из базы данных или документа) — и таким образом дополняет свои знания, что приводит к более точным и релевантным ответам.

Допустим, у твоей HR-команды есть пакет «Добро пожаловать в родительство», где описаны все льготы, положенные сотрудникам при пополнении в семье. С помощью LLM ты получишь креативное, не шаблонное письмо — но благодаря RAG оно еще и будет содержать реальные льготы, которые

предлагает компания. Это как если бы у твоего ИИ была «шпаргалка», в которую он заглядывает перед тем, как ответить (хотя да, он все еще может ошибиться — подробнее об этом позже).

Автоматизация ИТ

Мы специально отметили на рис. 4.1 момент перелома, потому что именно в этой точке компании не просто понимают, а начинают верить, что можно автоматизировать технологии и бизнес-процессы с помощью ИИ — и получить от этого серьезную отдачу. Это как раз один из примеров «горизонтальности» AI-сценариев. Подумай сам: любой крупный бизнес работает на технологиях — независимо от отрасли.

Мы сначала написали «любой бизнес», но это расстроило одну из наших мам — она немного опасается, что ИИ все захватит. Она сказала: «А вот у соседа — лимонадная лавка, и никакие технологии там не используются!» Правда? Джонни и Сьюжинка сами давили лимоны голыми руками? Ну ладно, допустим. (Честно говоря, мы вообще сомневаемся, что они использовали настоящие лимоны, но решили сохранить магию.) «А откуда у них лимоны?» — спросил автор. «Из супермаркета», — последовал ответ. А это уже цепочка: складская система, логистика, отслеживание — все на софте. На что мама сказала: «А если они сами выращивают лимоны? У меня в детстве было лимонное дерево». Впечатляет. Но прогноз погоды они смотрели? Постили ли что-то про свое лимонадное меню в TikTok? Принимали оплату через Venmo? Спор закончился, когда маму отвлек пуш от Netflix: ей порекомендовали «Адвоката Линкольна», потому что она поставила два лайка сериалу «Закон по Лидии Поэт». (Тут мы все согласны — оба отличные шоу.) Короче говоря: мы написали «крупный бизнес», чтобы не расстраивать мам, но на самом деле имелся в виду любой бизнес.

Теперь подумай, как в наши дни компании — большие или маленькие — управляют технологиями. В основном вручную, с опорой на армии технических специалистов, а в ряде случаев — на подрядчиков и консультантов. ИИ предлагает базовую технологию для автоматизации значительной части этой работы: обеспечение доступности систем, устранение критических уязвимостей, управление сертификатами, эксплуатация приложений и многое другое.

Может возникнуть вопрос: почему именно сейчас? Все просто — большая часть задач по управлению ИТ повторяющиеся: их нужно отслеживать, выполнять и контролировать. Оказалось, что ИИ отлично справляется с рутинной работой и, в ряде случаев, может устранять проблемы еще до того, как они появятся. Это и есть главное обещание ИТ-автоматизации.

Возьми, к примеру, скучную задачу — контроль состояния сертификатов. Централизованная система управления сертификатами кажется простой задачей, но для многих это реальная головная боль, потому что сертификаты раскиданы по распределенным системам. Из-за отсутствия общей видимости и ручного учета формируются фрагментированные реестры сертификатов, возрастает риск забытых и просроченных сертификатов. Часто они истекают до того, как кто-то об этом узнает, и это напрямую влияет на связанные приложения, вплоть до полной остановки сервисов.

Более того, большинство инструментов управления сертификатами не работают эффективно, потому что каждый из них видит только «свои» сертификаты, и трудно отследить, что именно сертификат стал причиной сбоя. Масштаб и последствия просроченных, устаревших или некорректно настроенных сертификатов колоссальны — это приводит к сбоям, рискам и нарушению непрерывности бизнеса. Просроченные сертификаты также открывают уязвимости для атак типа «человек посередине», когда злоумышленники могут перехватить конфиденциальные данные. Это именно тот тип задачи, который ИИ может решить эффективно.

Самое время вспомнить и о других вещах, которые ИТ-команды делают незаметно: работа с CDN,¹ системы обнаружения вторжений, управление электропитанием — и многое другое. Чтобы по-настоящему оценить как их вклад, так и потенциал ИИ в этих задачах, нужно понимать, что современные бизнес-приложения — это уже не простые программы. Это сложные экосистемы, построенные и управляемые с помощью множества инструментов и сервисов, развернутых в облаке и

¹ ChatGPT поясняет: CDN (Content Delivery Network) — это сеть распределенных серверов, которая позволяет быстрее доставлять контент пользователям за счет того, что данные (например, изображения, видео, веб-страницы) кэшируются и выдаются с ближайшего к пользователю сервера. Это снижает задержки, уменьшает нагрузку на основной сервер и повышает отказоустойчивость.

на локальной инфраструктуре. ИИ может помочь уже сейчас, но в будущем — это будет просто необходимостью.

Почему мы так считаем? Попробуй осознать следующий факт, которым поделился аналитик IDC: к 2028 году будет создано более одного миллиарда новых логических приложений. Это выходит за рамки этой книги, но однозначно говорит о том, что ИТ-директорам стоит изначально подходить к трансформации с гибридным мышлением. GenAI и гибридное облако объединяет одно — все крутится вокруг данных. Данные гибричны по расположению и типу. Гибридное мышление позволяет выжать максимум из данных, а значит — максимум из моделей. Вместе это усиливает бизнес и становится предпосылкой для продвинутой цифровой трансформации.

Короче говоря, по мере того как вы продвигаетесь в ИТ-автоматизации, это снижает технологические издержки. Ваша компания тоже попала на удочку, что публичное облако — это очевидная экономия, но по факту все вышло дороже? Используйте ИТ-автоматизацию, чтобы перенести эти рабочие нагрузки в более дешевую среду. Приложение не используется в полном объеме? Автоматизация позволит перераспределить неиспользуемые ресурсы. Возможности здесь практически безграничны — поэтому именно в этой точке начинается реальное создание ценности.

У IBM есть внутренняя инициатива под названием «Client Zero» — вызов от CEO добиться \$3 млрд экономии за счет повышения продуктивности через ИИ и превратить эти кейсы в шаблоны для клиентов. Это масштабный проект, потому что мы добиваемся реальных результатов, и делаем это с посылом: мы сами все это прошли. Но есть и другая сторона: IBM на этом пути тоже допускала ошибки, сталкивалась с неудачами — и именно они проложили дорогу к успехам. Эти уроки мы тоже передаем клиентам (в том числе и болезненные), чтобы они могли двигаться быстрее, увереннее и с меньшими потерями.

Эта книга не про IBM, и у нас нет возможности подробно рассматривать все проекты, которые помогли нам выполнить задачу, поставленную CEO. Но даже только результаты проекта по ИТ-автоматизации заслуживают внимания (помимо него были проекты с ИИ в HR, продажах, разработке, управлении объектами недвижимости и др.):

- 80% ключевых ИТ-проблем теперь решаются и локализируются ИИ
- Рост NPS (Net Promoter Score) в клиентской поддержке на 25 пунктов
- \$165 млн годовой экономии операционных расходов
- Снижение затрат на демонстрационную платформу TechZone на 66%
- Сокращение времени развертывания демонстрационной среды на 85%
- 45 000 автоматизированных действий с ресурсами в месяц

Пролистай этот список еще раз. Во-первых, важно понимать, что эти цифры, скорее всего, уже занижены на момент чтения книги — такие инициативы продолжают приносить результат. Во-вторых, среди этих пунктов наверняка есть тот, который бьет и по твоей боли — теперь у тебя есть шаблон, как подойти к решению.

Да, мы понимаем: автоматизация может казаться скучной. Это не то, что видит клиент напрямую. Но именно благодаря ей у тебя появляется время заниматься действительно ценным — разрабатывать продукты, лучше понимать потребности клиентов, находить новые пути для улучшения сервиса. Внутри компании автоматизация превращает тебя в героя. Мы однажды за завтраком обсуждали, насколько просто стало пользоваться AskIT — большинство ИТ-проблем решаются мгновенно или же тикет создается и закрывается с минимальными усилиями. Почему? Все верно: ИТ-автоматизация.

Код — язык компьютеров

ИИ меняет процесс творчества не только для писателей и художников, но и для разработчиков. Языки программирования постоянно эволюционируют, и разработчики всегда востребованы для адаптации приложений к новым стандартам. Это отличный пример использования ИИ, который неочевиден для тех, кто вне ИТ. Почему? Потому что, когда большинство людей слышит «язык», они думают о коммуникации между людьми (или с домашними животными). Но как обсуждалось в главе 2, если посмотреть на вещи с другой стороны — все становится языком. Код — это такой же язык, как и любой другой. Будь то COBOL, Java, Go или Python — так же, как люди бывают полиглотами, так и разработчики: они «разговаривают» с машинами на множестве языков. А машины — «понимают» разные языки программирования.

И по мере того как технологии и операции автоматизируются, высвобождаемые ресурсы можно направить на более ценные задачи — и одна из них всегда будет заключаться в создании нового.

Известно, что по поводу ИИ-помощников для кодинга идут жаркие споры между опытными и начинающими разработчиками. Оценки — самые разные. McKinsey, например, отмечает, что ИИ может сэкономить разработчикам 35–45% времени при написании кода (правда, только около 10% — при решении сложных задач). Но есть и скептики. Например, исследование кафедры компьютерных наук Университета Пердью, где проанализировали 517 вопросов с Stack Overflow, показало, что в 52% случаев ответы ChatGPT были неверными, а в 77% — избыточными по объему.

Более того, опытные программисты, такие как Джейсон Холл из Pirate Software, отмечают, что ИИ быстро генерирует код, но отладка этого кода потом занимает существенно больше времени, чем если бы код был написан человеком.

Команда из Purdue также провела лингвистический анализ 2000 случайных ответов ChatGPT и обнаружила, что они «более формальные и аналитические» и «менее негативны по тону». Если вы когда-либо искали помощь по программированию на публичных форумах, вы это оцените. Те, кто не пишет код, могут не знать, что для разработчика почти обряд посвящения — задать «глупый» (по мнению других) вопрос и получить жесткий ответ от кого-то, кто, возможно, писал на ассемблере в 15 лет. Пропустили что-то в ветке на сотню комментариев? Ждите RTFM (в нашем кругу — *прочти, наконец, руководство*, но есть и другие версии) и сердитый эмодзи. Иными словами, фраза «глупых вопросов не бывает» не особо котируется в среде девелоперов.

Мы уверены, что ИИ будет очень полезен программистам, но для этого нужен иной подход, отличающийся от того, как мир воспринимает ИИ-помощников по программированию сегодня. Во-первых, ИИ будет становиться лучше — это факт. Но важно задать вопрос: «На каких данных обучалась модель, помогающая писать код?» Это были просто фрагменты кода с Reddit, Stack Overflow, GitHub? Видели ли вы, откуда они взялись? Это принципиально важно и обсуждается в книге, особенно в главе 5. Много кода из этих источников раньше работало, но теперь — нет. Кто знает, тот поймет.

А что насчет специализированных моделей, обученных на проверенных и прозрачных данных? Например, watsonx Code Assistant for Ansible Lightspeed создан только для написания Ansible-плейбуков на YAML. Он обучен на сотнях тысяч кейсов Red Hat — ключевого разработчика Ansible и его корпоративной платформы Ansible Automation. Учитывая опыт Red Hat в Ansible-развертываниях, плюс данные из модерлируемого сообщества Ansible Galaxy, получается набор проверенных (работающих) и объяснимых (с понятным происхождением) данных для построения мощного ИИ-помощника.

Возвращаясь к инициативе IBM *Client Zero*, значительная часть эффективности в развертываниях была достигнута благодаря Ansible. С использованием watsonx Code Assistant for Ansible Lightspeed у разработчиков плейбуков уровень принятия сгенерированного кода составил 70%. Это огромный результат. Только на задачах по установке обновлений ОС IBM сократила время патча на 93%. Теперь понятно, почему IT-запросы обрабатываются быстрее — это позволило техкомандам сосредоточиться на более сложных задачах, а не тратить время на рутинные обновления, вроде патча CVE-2022-0847 (уязвимость Dirty Pipe в ядре Linux).

Есть еще один момент, к которому мы вернемся в главе 7, но стоит отметить уже сейчас: модель, лежащая в основе этого кодового помощника, в 33 раза меньше некоторых других моделей, которые вы, скорее всего, знаете в контексте генерации кода с помощью ИИ. Такая более точная и компактная модель (а значит, не нужны дорогие GPU, и ее можно запускать хоть на ноутбуке) — это тренд, который, как мы считаем, будет все активнее проявляться в будущем как в генеративном ИИ в целом, так и в ИИ-инструментах для программистов.

Итак, теперь, когда общие разговоры про генерацию кода ИИ позади, давайте подумаем о других аспектах разработки. Кодинг — это не только генерация. В том же исследовании McKinsey, на которое мы уже ссылались, отмечается, что помощники на базе GenAI используются также для рефакторинга, документирования и ответов на вопросы по коду.

Все мы в какой-то момент работали в лабораториях IBM и видели по-настоящему впечатляющий код — не только качественно написанный, но и отлично документированный. Но, честно говоря, мы также видели много того, что называем «спагетти-кодом»: все запутано (хотя код может быть рабочим), и

разобраться в нем — мучение, хуже, чем подавать налоговую декларацию. Поддерживать чистую, хорошо комментированную структуру кода действительно тяжело — по множеству причин: текучка кадров, работа с подрядчиками, уход людей на пенсию, временные назначения. Документирование часто оказывается самой малозначимой задачей для разработчиков, особенно тех, кто считает себя Моцартом программирования.

Ирония в том, что чем сложнее кодовая база, тем меньше она документирована. Автоматизация составления кратких описаний кода может освободить время программиста для более ценных задач — точно так же, как GenAI может создать первый драфт письма клиенту, он может подготовить черновой вариант документации. Конечно, это все еще требует участия разработчика, чтобы все проверить, но начальная стадия документации *сдвигается влево*, позволяя сосредоточиться на проектировании и создании нового функционала. А еще, по мере дообучения языковой модели, можно добиться единообразия в стиле и формате комментариев.

Ну и, конечно, у хорошо документированного кода есть эффект *вниз по цепочке*. Языковые модели позволяют взаимодействовать с "вещами", как контракты, законы и, да, кодовые базы. Представьте, что вы — новичок, только что нанятый и назначенный на проект. Вы хотите понять, как работает функция `processOrder()` и с чем она взаимодействует. Или выяснить, где еще в коде используется определенный участок. Просто спрашиваете ИИ, подключенного к кодовой базе — это сильно ускоряет адаптацию, сохраняя при этом принципы парного программирования.

Парное программирование в эпоху GenAI

Модель парного программирования — это совместная практика разработки ПО, где обычно объединяют джуниора (известного как «водитель», который пишет код) и сеньора («штурмана»), который проверяет код по мере его написания и анализирует его в контексте всего приложения. Традиционно они сидели за одним столом, но в условиях гибридной работы это стало сложнее. Такой подход предназначен для повышения качества кода, обмена знаниями и улучшения сотрудничества.

Однако эта модель может быть очень затратной по времени для сеньора, хотя и приносит огромную пользу. Сложный выбор. Что, если LLM, содержащая знания сеньора, возьмет на себя роль постоянного парного программиста для новичка? Сеньор останется доступным для еженедельных встреч, но ИИ возьмет на себя постоянное наблюдение за рабочим процессом. Более того, это позволяет реализовать современную стратегию локации работы — признавая, что талантливые разработчики могут находиться где угодно и не всегда могут делить физический стол или даже единое рабочее время в гибридной среде.

Другие факторы сделали документацию кода такой проблемой. В вихре Agile-спринтов разработчики, как кофеиновые хомяки, бегут в бесконечном цикле «код — тест — релиз». А поскольку документацию все равно никто не читает месяцами, она отодвигается на второй план. Но хорошая документация критически важна для любой команды. Она не только помогает другим разработчикам обновлять или проверять код, но и способствует обмену знаниями между командами. McKinsey обнаружил, что это область, где ИИ может помочь больше всего — сокращая время на 45–50%. IBM также нашла множество преимуществ в этом направлении, но результаты мы прибережем до главы 6, где расскажем о решении, появившемся в рамках IBM skills challenge.

Вы могли подумать, что ИИ автоматизирует разработку ПО. Действительно, создание ценности с помощью ИИ есть в генерации кода, автодополнении и создании документации, но это не вся работа. Новая разработка начинается с тезиса: какие приложения улучшат наш бизнес? Это удержание клиентов? Рост выручки? Улучшение сервиса? Как модернизировать legacy-приложения?

Эти мыслительные процессы ИИ выполнить не может, но создание решений для них можно усилить с помощью ИИ. Например, многие компании зависят от транзакционных систем и хотят модернизировать (или хотя бы понять) критически важные части своей кодовой базы. ИИ может помочь разобраться в старом коде, который плохо документирован, — а чаще всего никто и не представляет, как работает код, на котором держится бизнес! Эта «корпоративная амнезия», усугубляемая уходом опытных специалистов.

Чтобы что-то сделать с монолитными приложениями, их нужно понять, поэтому любой проект модернизации начинается с разгребания корпоративной амнезии. Это важный кейс использования

ИИ, который не связан с написанием кода. Как говорится, лучше один раз увидеть, чем сто раз услышать, поэтому мы нарисовали (пусть и не шедеврально), как ИИ может помочь:

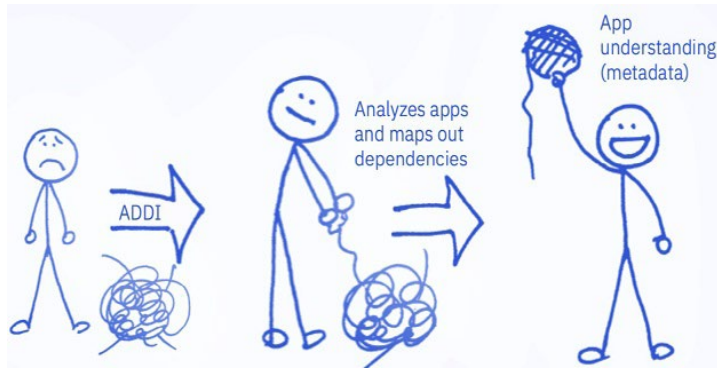


Рис. 4.2. Распутывание корпоративной амнезии

Слева показана ситуация, в которой сегодня находятся многие: куча кода, напоминающего запутанный клубок ниток. IBM столкнулась (и решила) эту проблему для многих своих клиентов, использующих мейнфреймы с крупными монолитными приложениями, где взаимозависимые функции годами обрастали изменениями. Подобный хаос можно найти и в других фреймворках — например, в Java, Hibernate, Spring и Enterprise Java Beans (EJB), у каждого свои «клубки».

Неважно, какой стек технологий вы используете: если у вас миллионы строк кода, сотни зависимостей и устаревшая документация, ваши разработчики могут потратить недели или месяцы, пытаясь понять, какие изменения нужны для модернизации блока кода перед его рефакторингом (и потенциальным переносом) в гибридное облако. И после всей этой работы нет гарантии, что их правки ничего не сломают — будь то из-за плохого кода или чего-то упущенного.

Хорошая новость: здесь тоже может помочь ИИ. Например, многие крупные компании до сих пор полагаются на COBOL-код. Этот код работает стабильно, но любые попытки его изменить или модернизировать напоминают игру в Дженгу — только падение башни грозит проблемами для бизнеса, а не просто рассыпавшимися деревянными блоками. ИИ помогает таким клиентам анализировать код, определять его функции, связи и многое другое.

IBM создала для этого ИИ-инструмент Application Discovery and Delivery Intelligence (ADDI). Как показано в центре рисунка — это инструмент анализа, который визуализирует приложения, данные и задачи на мейнфрейме. Он позволяет архитекторам и разработчикам одним кликом находить зависимости, вносить изменения и поддерживать документацию в актуальном состоянии. По сути, ADDI — это идеальный инструмент для распутывания клубков кода, помогающий обратить вспять корпоративную амнезию. Он работает как любопытный кот, разматывающий клубок: анализирует приложения и выстраивает карту зависимостей, превращая хаос в понятную структуру.

В правой части рисунка вы получаете понимание приложения (благодаря сгенерированным метаданным), которым можно гордиться как трофеем (главное — следить, чтобы кот не утащил свой клубок обратно).

Что написано на этом трофее? Уверенность! Теперь вы можете по-разному взглянуть на блоки своей «Дженги». Возможно, вы модернизируете часть приложения для новых сценариев использования — и теперь видите, что произойдет, если с этим блоком кода что-то пойдет не так, или какие бизнес-процессы от него зависят. Или у вас намечается волна уходов сотрудников на пенсию, и ваша задача — убедиться, что их код не останется без присмотра, обучив новых специалистов и отложив модернизацию на потом.

В реальности обычно все смешано: часть кода останется как есть, часть будет модернизирована по разным причинам. Главное — теперь у вас есть ясность для принятия этих решений.

Сравните это с тем, как некоторые клиенты подходят (по совету вендоров) к модернизации своего COBOL-кода с помощью GenAI. Прежде всего, они пропускают самый важный этап — понимание. Это создает ненужные риски — представьте, что вы начали капитальный ремонт дома без архитектурных чертежей, не зная, как устроен каркас и где несущие конструкции! Нам известны случаи, когда клиенты брали универсального AI-ассистента, чтобы на основе COBOL sourcebook'ов или фрагментов

кода автоматически сгенерировать Java-версию. Сделайте шаг назад и спросите себя: сколько высококачественных пар COBOL/JAVA-кода действительно доступно для обучения моделей (официально или неофициально)? Ответ: совсем немного. В итоге такие клиенты получали *JOBOL* — некачественный код, который толком не работал, был плохо задокументирован, трудно изменяем, и любые попытки доработать его, не сломав, занимали массу времени.

И, как и в случае с Ansible, IBM разработала watsonx Code Assistant for IBM Z для поддержки модернизации мейнфрейм-приложений. Но этот продукт не просто про миграцию кода. Он построен на платформе ADDI, которая охватывает весь жизненный цикл разработки (понимание, рефакторинг, трансформация, валидация, рекомендации, наблюдение). В продукт встроены механизмы автоматизации, которые сначала собирают метаданные и формируют понимание системы, затем выявляют бизнес-сервисы, которые можно выделить, и только потом — после анализа и рефакторинга — начинается миграция. Чувствуете разницу? Показательно, что многие клиенты используют этот инструмент исключительно ради этапа «понимание» — и это, пожалуй, лучшая похвала менеджерам продукта, которые действительно поняли боли своих клиентов.

А как именно эта система достигает такой эффективности там, где другие решения показывают слабые результаты? Ответ очевиден: у IBM есть программисты, владеющие и COBOL, и Java, которые создали тысячи пар функционально эквивалентных программ на обеих платформах. Специально для обучения watsonx Code Assistant for IBM Z использовались именно такие реальные, корпоративные пары COBOL-Java. Результат — высококачественный код без *JOBOL*.

AI в программировании — это только верхушка айсберга. Мы сравниваем AI-код-ассистентов с коробкой передач Tiptronic, которую Porsche представила в 1990 году (позже она эволюционировала в PDK — Porsche Doppelkupplung). Эта трансмиссия давала водителю выбор: получать удовольствие от ручного управления, использовать автоматическую коробку или находиться где-то посередине. Причем даже при ручном управлении, если водитель делал ошибку, которая могла повредить автомобиль, автоматика вмешивалась и предотвращала поломку.

Прекрасная аналогия, на наш взгляд. Вы — начинающий разработчик и хотите просто начать? Пожалуйста, вот вам сгенерированный код. Вы — эксперт, чьи коммиты скрипят от чистоты? Отлично, используйте AI, чтобы сэкономить на рутине или документировать тот божественный код, который вы только что написали. Вы только пришли в компанию и хотите быстро разобраться, как работает метод processOrder() и с чем он связан? Спросите у AI. Или вам поручили разобраться со старым, но критичным модулем (вроде системы расчета зарплаты), и вы не уверены, что именно может сломаться при переходе с COBOL на Java? В любом случае, относитесь к таким ассистентам не как к простым генераторам кода, а как к трансмиссии Porsche. Одно можно сказать точно: GenAI изменит опыт разработчика.

Цифровой труд и AI-ассистенты

Большинство людей с радостью наняли бы себе помощника: кого-то, кто выполнял бы задачи, которые им не хочется делать, которые повторяются или просто слишком сложны. Проблема в том, что нанять еще одного человека слишком дорого. Добро пожаловать в мир цифрового труда.

Раньше под термином «цифровой работник» понимали человека с цифровыми навыками. Сегодня это слово обозначает категорию программных роботов (не тех, что в фильмах — они тоже на подходе, но пока не здесь), обученных выполнять конкретные задачи или процессы в партнерстве с людьми. Их часто называют ботами. Мы определяем цифровой труд как программно реализованный труд, способный самостоятельно выполнять значимые части сложных сквозных процессов, используя широкий спектр навыков.

Цифровой труд может использовать AI для выполнения последовательностей задач в рамках заданного процесса. Цифровые сотрудники (боты) применяют возможности AI, такие как обработка естественного языка, агенты и генеративный AI (и другие), чтобы взаимодействовать и коммуницировать, мыслить и принимать решения, подбирать нужные навыки в реальном времени и применять их в контексте, сохраняя рабочую память о прошлых взаимодействиях.

Люди, работающие с информацией, могут обучать этих цифровых сотрудников, ставить им задачи и делегировать работу. Такая работа может варьироваться от автоматизации и ускорения простых задач до помощи в более сложных решениях. Например, цифровой сотрудник по расчетам с поставщиками

может автономно выполнять части четырех традиционных ролей — специалиста службы поддержки, агента по выставлению счетов, кассира и эксперта по урегулированию претензий — для выполнения процесса от заказа до получения денег (order to cash, OTC). Благодаря тому, что цифровые работники расширяют возможности своих человеческих коллег, их массово внедряют в рамках цифровой трансформации (shift left), что позволяет перераспределять персонал на более стратегические задачи (shift right).

Вы точно сталкивались с хорошими и плохими чат-ботами — одни действительно умеют вести себя как человек и выполнять полезные действия без ожидания на линии (это хорошо), другие с трудом справляются с двумя-тремя задачами (это плохо), а третьи настолько бесполезны, что напоминают бесконечное телефонное меню, в конце которого звучит: «Извините, мы закрыты» (это очень плохо).

В мире, где мгновенное удовлетворение стало нормой, потребители (а если честно, и ваши сотрудники) ожидают быстрых, если не моментальных, ответов на свои вопросы и проблемы. Бизнес понимает, что качественное обслуживание клиентов критично для долгосрочного успеха. К счастью, технологии не отстают от времени, и цифровая революция с AI помогают полностью переосмыслить работу колл-центров через цифровой труд.

AI-ассистент на базе цифрового труда — это универсальный кейс, потому что не существует бизнеса без фронт-офиса для клиентов или сотрудников (а если и существует — ему стоит его завести). Почему это так привлекательно? Вспомните, о чем мы говорили ранее: если человек поднимает трубку и обслуживает звонок, это стоит около \$5 за простой кейс; если тот же звонок обрабатывает цифровой работник, это обходится примерно в \$0.25.

Например, Департамент автотранспорта штата Орегон внедрил цифровой труд, чтобы справиться с резким ростом объема входящих звонков во время пандемии COVID-19. Всего за несколько недель они начали автоматически обрабатывать около 30% базовых запросов. Эти цифровые сотрудники освободили время для живых сотрудников, чтобы те могли заниматься более сложными обращениями, и помогли сократить время ожидания клиентов. Эта инициатива сэкономила почти \$3 млн (двухлетняя стоимость около 30 сотрудников) и сократила время ожидания клиентов. Впечатляет!

Еще один пример shift left в пандемию — компания CVS, один из лидеров в сфере здравоохранения в США. Хотя у CVS множество каналов предоставления медицинских услуг — от страхования до программ благополучия, компания более всего известна своей сетью аптек. До пандемии у CVS было около 40 000 магазинов, и она обрабатывала в среднем около 150 000 звонков в день. Всплеск вызовов во время COVID заставил их срочно искать решение. CVS внедрила цифровой труд с применением AI, чтобы справиться с миллионами звонков. Благодаря значительным двузначным показателям автоматической обработки, CVS смогла обрабатывать более чем в четыре раза больший объем звонков по сравнению с допандемийным уровнем, ежедневно экономя огромные суммы.

Мы делимся этими двумя примерами не только потому, что хорошо их знаем (они были реализованы на технологиях IBM), но и потому, что они показывают: неважно, работаете ли вы в небольшом государственном агентстве или в крупной корпорации — цифровой труд может работать и на вас. А поскольку это примеры времен пандемии, это означает, что преимущества AI были доступны уже какое-то время. Новое поколение AI-агентов выведет такие решения на следующий уровень. Вопрос теперь в следующем: а что делает ваша компания?

Camping World — американская компания, продающая дома на колесах (RV) и комплектующие к ним. Она хотела освободить больше времени для своих операторов, чтобы те могли вести действительно содержательные и полезные диалоги с клиентами. Это означало — убрать из очереди к живым агентам простые и быстрые запросы, которые можно автоматизировать. Camping World внедрила технологию цифрового труда IBM, и ее сотрудники достигли двузначных приростов эффективности, а время ожидания на линии для клиентов, которым действительно нужен был живой оператор, сократилось до 33 секунд! И хотя у клиентов почти не оставалось времени, чтобы задуматься о жизни (они могут наверстать это на отдыхе в RV), удовлетворенность клиентов выросла — индекс вовлеченности у Camping World увеличился на 40%.

Еще один пример — сеть парикмахерских Sport Clips. У нее необычный подход: в каждом салоне телевизоры показывают всевозможные спортивные матчи, пока вам делают стрижку в стиле MVP. У Sport Clips масштабный план по франчайзинговому расширению, включая рост персонала на 30%,

чтобы достичь бизнес-целей. Компания поделилась проблемой: как франчайзи смогут найти нужных сотрудников, чтобы обслуживать клиентов и обеспечивать фирменный «чемпионский» уровень качества? Команда Sport Clips хотела дать франчайзи уверенность, что, присоединившись, они смогут найти мастеров, способных справляться с любой шевелюрой и делать каждую стрижку выдающейся. В итоге Sport Clips использовала продукт watsonx Orchestrate для создания цифровых работников и сократила процесс привлечения кандидатов с трех часов до трех минут.

Очевидно, для этого всего не обязательно использовать технологии IBM. Не поймите нас неправильно: мы были бы рады, если бы вы выбрали именно их, и мы искренне считаем, что GenAI-платформа IBM охватывает все аспекты, о которых вам стоит задуматься. Но, в качестве не-IBM-примера, компания Klarna объявила в начале 2024 года, что заключила партнерство с конкурентом IBM для внедрения цифрового труда в свою службу поддержки клиентов; сегодня ИИ обрабатывает две трети обращений (что сопоставимо с показателями CVS), что привело к снижению повторных обращений на 25% и эквивалентно работе 700 штатных сотрудников, благодаря чему персонал смог сосредоточиться на более сложных задачах. Это серьезный сдвиг вправо!

Видите закономерность? От найма парикмахеров до кемпинга на автофургонах, от получения информации по рецепту до оформления водительских прав для управления этим автофургоном — и во всех промежуточных точках — цифровой труд может помочь любому бизнесу, в почти любой задаче, в любой отрасли. Именно поэтому мы так настойчиво говорим о горизонтальном подходе к сценариям использования GenAI. Каждой компании может понадобиться помощь в найме, обслуживании клиентов, предоставлении информации и многом другом. Именно поэтому мы рекомендовали вам еще в Главе 1 разложить бизнес-процессы на наименьшие составные части в рамках совета по сценарию Dimension One — это дает исключительную ясность в том, где можно создать цифровых сотрудников для помощи людям в рабочем процессе и приблизиться к состоянию AI+.

Представьте, что каждый сотрудник в вашей организации был бы дополнен 10 цифровыми работниками. А почему бы и не 100? В Главе 1 мы говорили: сначала «сдвиг влево», чтобы затем можно было «сдвинуться вправо». Речь идет о построении автоматизации на ранних этапах рабочего процесса или «работы, которую нужно выполнить». ИИ прекрасно справляется с этим — если мы потратим время, чтобы объяснить, что мы хотим от него. Другими словами, это и есть создание цифрового сотрудника.

Разумеется, вопрос сохранения рабочих мест всегда всплывает, когда речь заходит о цифровом труде. Позвольте предложить вам другой взгляд на эту проблему. Начнем с фактов, изложенных в главе 3: большинство стран сталкивается со снижением численности населения (будущий или текущий дефицит рабочей силы) и падением производительности. Задайте себе вопрос: что, если Великая отставка на самом деле была Великим апгрейдом — возможностью привлекать и удерживать сотрудников, лучше используя их навыки? Разве цифровой труд не сделает это возможным, взяв на себя рутинную работу ваших сотрудников? Сотрудничая с ИИ, вы можете освободить своих работников от утомительных, малозначимых задач, позволяя им сосредоточиться на работе, для которой их нанимали. Вместо замены сотрудников это ставит их во главе процесса. И вместо увеличения затрат это оптимизирует ваш бюджет. Именно поэтому мы считаем, что главная ценность ИИ будет заключаться в предотвращении найма на роли, не создающие ценности, и в повышении привлекательности вашей компании как места работы.

Агенты

Вы, вероятно, уже догадались об этом в предыдущем разделе, но агенты — это следующий уровень продуктивности для бизнеса. ИИ-агент — это система или программа, способная выполнять задачи, проектируя собственный рабочий процесс и используя доступные инструменты (например, возможность сканировать веб-сайт, запускать код на Python, выполнять SQL-запросы, узнавать погоду и многое другое). Проще говоря, агенты расширяют круг задач, которые можно решить с помощью вашей GenAI-трансформации, потому что они не просто синтезируют информацию, но и могут разрабатывать собственный план решения поставленной задачи, выполнять действия и запоминать информацию! Правильная LLM в рамках агентного рабочего процесса творит чудеса, потому что агенты могут действовать автономно, итеративно прорабатывать проблемные области, адаптироваться, рассуждать и многое другое. Мы подробно рассмотрим агентов в главе 7, а в этом разделе поделимся лишь несколькими примерами использования.

Мы уже говорили о впечатляющих результатах цифрового труда в создании беспрепятственного клиентского опыта в предыдущей части нашего обсуждения ИИ, но ИИ-агенты могут вывести все это на новый уровень. ИИ-агенты могут быть интегрированы в веб-сайты и приложения, чтобы улучшить взаимодействие, наделяя виртуальных помощников еще большими возможностями. Например, представьте, что вы хотите заняться определенным видом спорта в определенном месте для идеального отпуска. Это запрос, который требует большего, чем просто указание дат на сайте бронирования. Агент может буквально составить маршрут (места для посещения, развлечения, еда) вместе с лучшим временем для поездки, чтобы максимизировать шансы на идеальную погоду для вашего вида спорта, минимизировать туристическую нагрузку, начать процесс оформления визы и даже забронировать билеты и столик в том замечательном ресторане, который знают только местные.

А если вы управляете государственным агентством, которое помогает выпускникам федеральной программы повышения квалификации найти новую работу? Агент может изучить вакансию, на которую претендует кандидат, понять самого кандидата (возможно, он из когорты обескураженных работников, готовых попробовать снова, только что закончил военную службу или, может быть, давно не проходил собеседования, потому что долго работал в компании, которая недавно закрылась), исследовать компанию, разместившую вакансию, в интернете и провести имитационные собеседования!

Задумайтесь о преимуществах. Кандидат не только будет лучше подготовлен к собеседованию (ведь второго шанса произвести первое впечатление не будет), но и, скорее всего, покажет лучшие результаты, потому что его уровень тревожности снизится после нескольких итераций с цифровым интервьюером. Можно добавить реальный голос и аватар, работающие на нескольких языках, что поможет кандидатам, для которых английский не является родным, улучшить другие аспекты самопрезентации. Возможности бесконечны.

Или рассмотрите реагирование на стихийное бедствие. Агентный ИИ может использовать LLM и инструменты (такие как данные о трафике в реальном времени, информация о локализации серьезных повреждений или сообщениях в соцсетях, связанных с бедствием, погода и т. д.) для выявления потенциальных сигналов бедствия от людей, вероятно находящихся в наиболее пострадавших районах, через их посты в соцсетях. Местоположение этих пользователей можно нанести на карту и объединить с данными о погоде, трафике, загруженности больниц и другими атрибутами, чтобы спасательные службы помогли большему числу людей за меньшее время.

Агенты могут также работать на разработчиков, автоматически создавая тестовую среду для выполнения кода в изолированном окружении, анализируя логи, генерируя обратную связь по коду и тестовому запуску, а возможно, даже предлагая советы по улучшению кода. Конечно, большую часть этого можно сделать асинхронно с помощью хорошего кодового ассистента с элементами GenAI. Но это очень итеративный процесс, и не будет никакого корпоративного знания, применяемого к другим, пишущим похожий код, чтобы развивать (или внедрять) лучшие практики компании. (Мы упоминали, что агенты могут учиться со временем? Подробнее об этом в главе 7.)

Если вы следуете советам из этой книги, то именно здесь вы буквально парите, потому что строите свою стратегию AI с моделями, которым доверяете, которые можно объяснить и которые управляются вашими данными. Агенты затем приносят масштаб в ваши усилия по GenAI, чтобы сделать еще больше работы за вас. Наконец, стоит отметить, что поскольку агенты могут действовать автономно, они требуют еще большей наблюдаемости и контроля, и поэтому мы описали (и еще раз обсудим позже) в этой книге, почему вам нужно твердое управление вашим AI.

Бизнес-подход: кейсы использования в горизонтальной плоскости

Когда мы переходим к бизнес-взгляду, природа кейсов использования меняется. Вы, вероятно, заметили это в предыдущем разделе, но в нашей работе есть три неоспоримых топовых кейса использования AI. Давайте назовем их «Большой тройкой»: обслуживание клиентов, кодирование, цифровой труд.

В главе 1 мы говорили о расходовании денег, чтобы сэкономить, и о расходовании денег, чтобы заработать, как об одном из видов классификации кейсов использования. Без сомнения, это звучит так же замечательно на TED-конференции, как и в этой книге, но теперь пришло время превратить этот слоган, который полюбит ваша руководящая команда, в реальную работу. Если вы только начинаете, «Большая тройка» — это те кейсы, которые стоит привязать к этой структуре.

Как только бизнес решит вопросы с «Большой тройкой», подумайте, куда еще может быть применен AI и огромный спектр кейсов использования, которые можно исследовать в любой из этих функциональных областей: технологические операции, цепочка поставок, закупки, финансы, человеческие ресурсы, маркетинг, юридический отдел.

Мы считаем, что 90% создания ценности в следующее десятилетие будет связано с одной из этих 10 областей (Большая тройка и семь функциональных областей выше).

Бонусный (горизонтальный) кейс использования — синтетические данные

Сейчас много говорят о высокой потребности в дорогих GPU для создания более мощных LLM, но есть одна тема, которой уделяют недостаточно внимания — то, что мы называем «засухой данных». Поскольку большая часть интернета уже была сжата в рамках типичной LLM, осталось очень мало данных, которые могли бы привести к новым прорывам (а LLM просто обожают данные). Отложим это в сторону и поговорим о таком типе засухи данных, который касается нас напрямую: представьте, что вы — компания, выпускающая кредитные карты, и вы хотите использовать GenAI для предотвращения мошенничества.

У вас, вероятно, есть масса данных о транзакциях, но сколько у вас данных о мошенничестве? Конечно, какие-то данные есть. А 50% ваших данных — это мошеннические операции? Если бы так, вы, скорее всего, уже были бы банкротами. А как насчет данных, попадающих под категорию персонально идентифицируемой информации или нечто подобное? Насколько проще было бы направлять модель с помощью данных, которые семантически похожи на ваши, но при этом вымышлены — это устранило бы кучу проблем с конфиденциальностью. И как подсказывали Саймон и Гарфанкел в своем хите «Fakin' It», ответом на засуху данных являются синтетические данные — другими словами, притворяйся, пока не добьешься.

Синтетические данные — это искусственно созданные данные, призванные заменить (возможно, из-за конфиденциальности) или дополнить (если настоящих данных недостаточно) реальные данные. Эти новые данные могут использоваться как наборы для тестирования и все чаще применяются для обучения моделей благодаря преимуществам в области конфиденциальности данных.

Синтетические данные могут также использоваться в здравоохранении — для защиты данных пациентов и повышения эффективности клинических испытаний при соблюдении нормативных требований. Другие примеры включают создание симулированных упражнений по проверке соблюдения нормативов для аудиторов, генерацию данных-плацебо, моделирование рисков для стресс-тестирования финансовых сценариев, краш-тесты, имитации переписей населения и многое другое.

Данные и искусственные, синтетические данные отражают реальные события на математической и статистической основе. Ранее Gartner предсказал, что к концу 2024 года 60% данных, используемых для обучения AI-моделей, будут синтетическими. С тех пор, как появился этот прогноз, многое изменилось — включая выпуск ChatGPT и то, что стало для GenAI аналогом момента Netscape. И хотя мы не знаем точную цифру, мы знаем одно — каким бы ни было это число, с каждым годом оно будет только расти (именно поэтому мы подробно объясняем, почему синтетические данные критически важны для обучения LLM, в главе 7).

Наконец, так же как реальные данные в моделях могут отражать предвзятость, синтетическая генерация данных, имитирующих естественные, вызывает аналогичные опасения, которые необходимо учитывать. По мере того как мы вступаем в этот вымышленный мир данных, становится ясно: возможно, этот кейс не привлечет к вам тусовщиков, но он определенно горяч, горяч, горяч — и поиск баланса между синтетическими и реальными данными будет формировать саму суть кейсов использования GenAI в будущем.

Набор кейсов использования по вертикали

Мы не просто намекали — мы уже прямо заявляли, и повторим еще раз: многие кейсы использования GenAI и агентных систем сводятся к горизонтальному применению. Такой подход даст вам прочную основу для выбора AI-кейсов. В конце концов, AI, определяющий изменение степени поглощения света родинкой для оценки риска меланомы, работает точно так же, как и AI, определяющий пористость или коррозию в свойствах материалов на выходе производственной линии. Видите? Это

два кейса из двух отраслей (здравоохранение и производство), но с одинаковой технологической основой.

Тем не менее, как только вы освоите горизонтальное мышление и поймете, что AI может делать в любой отрасли, следующим шагом будет «повернуть» это мышление на 90 градусов и взглянуть на применение AI в контексте именно вашей отрасли — по вертикали. В конце концов, убедить топ-менеджмент внедрить AI в автомобили доставки для расчета динамических коэффициентов внимания водителей (обратная метрика по отношению к степени отвлеченности) с целью снижения страховых взносов, ссылаясь на то, как AI распознает картошку с бородавками на линии мойки, — может не прозвучать достаточно убедительно. (К слову, теперь у вас уже есть четыре кейса из разных отраслей с применением одной и той же горизонтальной AI-технологии.)

В этом разделе мы рассмотрим несколько отраслей и приведем конкретные примеры, а также список (без пояснений) различных AI-кейсов, которые вы сможете изучить самостоятельно. Хотя мы и называем эти отрасли вертикалями, на деле вы можете воспринимать их и как роли. Например, в сети больниц будут отделы бухгалтерии и юридической поддержки, как и в большинстве других компаний. Именно поэтому мы рекомендуем прочитать все кейсы, даже если они напрямую не связаны с вашей отраслью. И самое главное — помните, что мы говорили в начале этой главы: кейсы использования находятся всего в одном поисковом запросе от вас.

Сельское хозяйство

Это трудоемкая отрасль, но она также богата технологиями: от уборочных машин с управлением по спутниковым данным до аналитики фондового рынка, позволяющей оптимизировать время сбора урожая. Автоматизация ключевых процессов продолжит вносить инновации, способствующие повышению урожайности, улучшению состояния полей и снижению зависимости отрасли от высокорисковых факторов, таких как вода и вредители.

AI может помочь во многих аспектах. Например, в мире царит избыток пестицидов, потому что мы часто применяем метод «опрыскай и надейся» для борьбы с вредителями, патогенами и болезнями, способными уничтожить урожай. Оказывается, AI может сделать наземную сельхозтехнику умнее и эффективнее. Точное опрыскивание с поддержкой AI помогает фермерам обрабатывать только те части растений, которым действительно требуется защита от конкретной угрозы. Такой подход существенно снижает зависимость от широкого спектра гербицидов, уничтожающих все подряд.

AI переводит подход от «опрыскай и надейся» к «увидел — обработай». Даже состав удобрений и пестицидов может быть точно адаптирован под состав почвы или характер угрозы. В одном из решений AI способен распознавать 26 заболеваний сельхозкультур (например, кедрово-яблочную ржавчину и другие грибки) и за 0,2 секунды наносить защитный препарат с точностью до четверти дюйма. Фактически, в США уже работают роботы для сбора салата с поддержкой AI, которые обрабатывают более 10% урожая салата в стране и при этом позволяют сократить использование пестицидов на этой культуре до 90%!

Другие кейсы: животноводство (например, Lowisa и их лазерная терапия с AI для лошадей), автономный сбор урожая, агрономия, севооборот, отслеживание растительности с дронов, безопасность сбора урожая, оценка состояния и поведения скота, носимые устройства для животных, интеллектуальное орошение, вмешательство в кризисных ситуациях психического здоровья, определение качества и сортировка продукции, безопасность монокультуры, селекция семян и размещение по полю, смешивание семян, прогнозирование плодородия почв, управление затратами, устойчивое повышение производительности, прогнозирование погоды и многое другое.

Бухгалтерия

Технологии кардинально меняют сферу профессиональных услуг, и бухгалтерия — одна из тех областей, где перемены будут особенно глубокими. GenAI и агентные системы трансформируют работу налоговых и бухгалтерских специалистов и будут все дальше отдалять эту профессию от подготовительных задач (которые в значительной степени будут автоматизированы AI) в сторону консалтинга и аудита. Если вы бухгалтер — это хорошие новости. Во-первых, профессия никуда не исчезнет — калькулятор ведь не убил бухгалтеров, правда? Представьте себе бухгалтера без калькулятора! Вас ждет серьезный рост доходов, если вы пойдете по тому пути, который описан в Главе 6. Это поможет избавиться от образа «счетоведа» — вы станете похожи на повара, который перестал чистить картошку и начал разрабатывать меню на звезды Мишлен.

Подготовительные услуги, которые традиционно были основной занятостью бухгалтеров, становятся все более стандартизированными, что в свою очередь снижает ставку оплаты за час работы. По мере того как AI берет на себя все больше задач, устаревший (но все еще используемый в юридических фирмах) подход, при котором стажеры выполняют задачи по почти партнерским тарифам, теряет актуальность.

Однако именно бухгалтеры всегда были надежными хранителями и контролерами финансовой информации — кто еще может лучше всех возглавить инициативу по управлению AI и оценить, в каких бизнес-подразделениях его можно эффективно применять с учетом финансовой целесообразности? Кто лучше поможет привязать потенциальную рентабельность инвестиций (ROI) к предлагаемым кейсам? Когда подготовительные услуги «смещаются влево», бухгалтеры могут сосредоточиться на том, чтобы помогать бизнесу в его развитии.

И это касается не только тех, кто работает с налогами и балансами. Близкие по духу профессии — управление казначейством, аудит, расчеты с поставщиками, оптимизация процессов «от заказа до оплаты», оценка кредитных рисков. Все эти области охватывают множество финансовых процессов, которые помогают оптимизировать и контролировать денежный поток, ликвидность, источники финансирования и соблюдение требований.

Например, AI поможет бухгалтерам лучше оценивать и расставлять приоритеты для ключевых факторов в *pro forma* (это бухгалтерский термин для проектируемой финансовой отчетности), чтобы формировать отчетность и влиять на прогнозируемые финансовые результаты. Путем консультирования руководства по этим данным, бизнес получает возможность принимать решения, основанные на данных, и улучшать точность прогнозирования.

Это полезно для сотрудников казначейства. По данным Treasury Today Group, только 20% компаний надежно прогнозируют денежную позицию на месяц вперед; при горизонте в три месяца эта цифра снижается до 5%. AI поможет компаниям минимизировать избыточную ликвидность и эффективно использовать излишки (например, для инвестиций, капитальных расходов и т. д.).

Другие кейсы: автоматический аудит на соответствие, дебиторская задолженность, прогнозирование ликвидности, оценка контрагентов, объяснение сложного налогового законодательства для неспециалистов, предотвращение и выявление мошенничества, автоматическая генерация финансовой отчетности, управление и оптимизация ликвидности, операционные и репутационные риски, моделирование аудиторских проверок, управление затратами, анализ налогового разрыва, оценка последствий изменений в налоговом кодексе и их интерпретация и многое другое.

Образование

Из всех отраслей образовательные учреждения, возможно, сталкиваются с самой большой трансформацией из-за AI — мы говорим о том же уровне, что и индустрия каршеринга и такси. Педагогам нужно переосмыслить, что значит учиться, оценивать и развиваться в среде, где студенты могут использовать AI для выполнения значительной части своей работы. Подумайте об этом: как дать студенту задание по программированию, если он может использовать LLM, чтобы написать код? А как насчет эссе?

Это очень спорная тема сейчас, и нам жаль, что мнения о применении GenAI и агентов в классе так разнятся. Чтобы сказать прямо, это не обсуждение «следует ли это использовать». Многие вели такие же дебаты о калькуляторах. Время пришло. Существует множество креативных способов внедрения AI в класс. Например, что если команда использовала бы AI для создания трехстраничного резюме по теме, а студенты собирались бы вместе, чтобы «улучшить» язык? В конце концов, это то, что делают юристы сегодня — разве это не научит их тому, как будет выглядеть будущая работа? А как насчет критики того, что написал AI? Проверки и цитирования источников на предмет достоверности (это научит студентов правильно использовать AI из-за его склонности к «галлюцинациям»).

Учителя также могут переосмыслить уроки, где AI может помочь. Как насчет использования LLM для создания сложных ролевых разговоров? Студенты могли бы вести дебаты с AI, принявшим облик исторической фигуры, и «путешествовать во времени», чтобы поспорить с ней. Будучи студентами, мы бы нашли интересным поговорить с бывшим канадским премьер-министром, который в 1863 году, после того как его стошнило во время дебатов (он был пьян!), заметил: «Меня иногда тошнит не из-за выпивки или каких-либо других причин, кроме как из-за того, что мне приходится слушать бред моего

почтенного оппонента». Возможно, не с той исторической фигурой, которую мы привели в примере, но представьте себе такое взаимодействие с AI-генерированным персонажем в его облике, с голосом и всем прочим.

Суть в том, что педагоги могут использовать LLM, чтобы вовлекать учащихся и переключать фокус на искусство критического мышления, а не на заучивание фактов. Чтобы сказать прямо, джинн выпущен из бутылки, и если некоторые педагоги продолжают идти своим собственным ритмом, они, вероятно, отстанут от реального ритма и в конечном итоге подведут тех самых людей, которым пытаются помочь. Здесь нужно найти общий язык — и это нужно сделать быстро и в соответствии с миром, в который студенты фактически вступают.

Одно мы знаем точно: педагогам нужно повышать квалификацию, чтобы не избегать, а стремиться внедрять AI в класс. Иронично, но AI может помочь учителям бесчисленными способами — от снижения нагрузки по оценке и планированию уроков до управления растущими соотношениями учеников и учителей и многого другого.

Поэтому мы написали главу 6, которая полностью посвящена повышению квалификации. Это не будет легко. В образовании редко случаются моменты сдвига. Но вы не можете подготовить детей к работе будущего, если не знаете ничего о навыках будущего. Честно говоря, будущее образования будет все меньше и меньше связано с подготовкой студентов к конкретной работе. Мы думаем, что это будет подготовка ко «всему и чему угодно», что принесет будущее. Нравится вам это или нет, студенты выпустятся в мир, полный AI, и им нужно будет знать, как ориентироваться в этих инструментах, понимая свои возможности и ограничения.

Мы считаем, что тестирование запоминания фактов является проблемой как в школах, так и за их пределами. Не поймите нас неправильно; у него есть свое место, но оно стало стандартом по умолчанию, потому что так было всегда, а не потому, что это работает. Это всегда нас беспокоило — и не потому, что мы ошибались в вопросах на сертификационном экзамене администратора баз данных, где нужно было вспомнить точное ключевое слово для изменения размера предварительной выборки для буферного пула из списка сложных синтаксисов с тонкими различиями, которые никому не интересны. А как насчет вопросов «выберите все подходящие», которые не указывают, сколько вариантов правильных? Даже не будем об этом.

Представьте себе экзамен по базам данных, который создает живую среду с условием переполнения буферного пула, где кандидат должен справиться с этой ситуацией. Мы не говорим о контролируемых проверках на вашем ноутбуке, а о свободном доступе к интернету и LLM. Так люди работают сегодня. Они учатся, исследуют, учатся, и так далее. Это еще не здесь, но мир должен двигаться дальше от искусства запоминания для многих аккредитаций, где оно используется сегодня.

Отличный пример использования GenAI и агентов — это раннее выявление уязвимых студентов и получение более точных данных о причинах отчисления. Ведь студент обычно бросает учебу из-за сочетания обстоятельств, которые не всегда связаны с академической успеваемостью. AI можно использовать, чтобы удерживать больше студентов, используя его прогностические элементы для своевременного вмешательства, пока не стало слишком поздно.

Агенты также могут помочь и после выявления проблем со студентом. Сообщения, бронирование встреч, уведомления для более широкой поддерживающей экосистемы (профессор, ассистенты, регистратура, бухгалтерия и др.). Но прежде чем AI сможет помочь, вам нужны данные (вся эта штука с IA); помните, чтобы решить любую сложную проблему, нужно сначала собрать точки, прежде чем их можно будет соединить.

Есть также множество примеров использования образовательных учреждений, которые не имеют ничего общего с образованием. Вы найдете их в других разделах этой главы, но если учесть, что примерно 70% затрат университетов связаны с персоналом, здесь можно помочь больше, чем кажется на первый взгляд.

Другие примеры: изменение среднего диапазона оценок, обучение и тестирование с использованием дополненной или виртуальной реальности, разработка учебных программ, динамическое тестирование, помощь в изучении языка, управление вмешательствами, операционный и риск-менеджмент, оптимизация проживания, выявление отстающих, удержание студентов, планирование оплаты обучения и многое другое.

Здравоохранение

Это сложный вопрос, потому что текущие темпы роста затрат на здравоохранение для большинства стран просто неустойчивы. Например, уровень инфляции в США с 2011 по 2019 год составил 1,75%. Для здравоохранения он составил 7,4% — это более чем в четыре раза больше! Добавьте к этому увеличение числа хронических заболеваний (ожирение, депрессия, диабет и теперь одиночество) и старение населения, и вы быстро поймете, что нужно что-то делать срочно.

Когда вы последний раз видели терапевта? Получили ли вы хотя бы 15 минут прямого зрительного контакта с ним? Дело не в том, что они вас не любят, но большая часть их времени уходит на ввод данных о вас в компьютер, а затем на переход к следующему пациенту в их перегруженном расписании. Это потому, что медицинские работники отмечают, что тратят почти 40 часов в неделю на клиническую документацию, общение и согласования со страховыми компаниями; фактически, время, потраченное на различные административные задачи, является одной из главных причин выгорания сотрудников, что, конечно, приводит к нехватке персонала.

GenAI и агенты имеют здесь такой большой потенциал, что мы чуть не исключили этот раздел, потому что не чувствовали, что сможем воздать ему должное. Тем не менее, у GenAI также сложный путь впереди между регулированием, выбором приоритетных случаев использования и этикой.

Прогнозирование — это огромный случай использования в этой области. Очевидный пример — прогнозирование начала заболевания, но с увеличением нагрузки на доступность клиницистов прогнозирование нагрузки на пациентов и оптимизация штата на основе этого прогноза являются ключами к снижению стресса в системе и требуют немедленного внимания. Например, одна больница в Торонто использует один алгоритм AI для большинства расписаний медсестер в их отделении неотложной помощи. Та же больница может буквально предсказать, что в субботу, между 13:00 и 18:00, в их отделении неотложной помощи будет около 80 пациентов, из которых 10 будут иметь проблемы с психическим здоровьем, а 12 будет трудно лечить.

Способность AI обрабатывать невероятное количество информации также дает шанс изменить ситуацию с редкими генетическими заболеваниями, которые приносят разрушительные истории страданий своим жертвам в течение многих лет. Они переживают эмоциональные страдания, посещая десятки, а то и сотни врачей в отчаянных попытках выяснить, что не так, но так и остаются без ответов. Текущая система направлена на выявление таких групп, которые упускаются системой.

AI может изменить шансы здесь и вывести генетические диагностические оценки на передний план медицинской помощи, а не оставлять их в конце, где они находятся сейчас. Наконец, в том, что может стать самым значительным потенциалом для спасения жизней, новое канадское исследование показало, что приложение AI для раннего предупреждения помогло одной больнице предотвратить неожиданные смерти на 26%. Можно ли масштабировать это более широко и повторять? Мы не уверены, но это внушает надежду.

Не забывайте, что вышеупомянутые (и последующие) примеры использования не предназначены для замены медицинского персонала, а скорее для того, чтобы служить дополнительным инструментом для ухода за пациентами и администрирования учреждений, чтобы улучшить результаты для пациентов.

Готовы ли вы исследовать роль AI в современной медицине? Возможно, начните со следующих примеров использования (и не забудьте те, которые мы рассмотрели в главе 1): поддержка послеоперационного ухода, клинический анализ, обобщение данных или документация (любой вид клинического рабочего процесса), предотвращение выгорания врачей и медсестер, сводки при выписке, раннее выявление заболеваний, улучшенное планирование питания, выявление мошеннических медицинских кодов, проверка медицинских кодов, интеграция электронных медицинских карт, снижение медицинских ошибок, поддержка памяти, предварительная оценка МРТ и других сканирований, соблюдение предписаний, обработка платежей, общение с пациентами, сообщения через портал, регистрация и обобщение, заметки врачей, помощь в радиологии, коммуникация по расписанию и многое другое.

Страхование

Страхование вращается вокруг понимания и страхования рисков, и сегодня страховым компаниям приходится иметь дело с очень сложными рисками. Рассмотрим страхование имущества и

ответственности (P&C) и влияние все более частых и разрушительных природных катастроф. Это критическая необходимость. Ведь минимальная оценка ущерба от пожара в Калифорнии в 2025 году составляет 150 миллиардов долларов, а максимальная — до 250 миллиардов. А как насчет риска при ценообразовании полисов страхования жизни с растущим числом нездорового населения и новыми рисками, такими как пандемии? Мы уверены в одном: страховая отрасль является основным кандидатом на изменения благодаря GenAI и агентам, и это окажет значительное влияние на страхование — в частности, на урегулирование убытков, распределение, подписание, оплату по мере использования (мы привели отличный пример в главе 1), персонализацию (переосмысление отношений между страховщиком и застрахованным) и ценообразование рисков.

Страховые компании могли бы использовать GenAI в рабочем процессе агентов с вызовами инструментов, чтобы укрепить лояльность среди страхователей, помогая им избегать потенциальных убытков, связанных с погодными условиями. Представьте себе: один град в Фениксе однажды нанес ущерб на 20 миллионов долларов! Оценочно, погодные оповещения дают страховщикам P&C возможность связаться со своими клиентами 10 раз в год (а не только при продлении полиса или при наступлении страхового случая). Как владелец автомобиля на пути града, вы не хотите, чтобы ваша машина была повреждена. Да, страховка покроет ущерб, но это хлопотно, есть риск повышения тарифов в будущем и так далее, поэтому мы предполагаем, что вы предпочли бы избежать такой ситуации. И мы уверены, что ваша страховая компания тоже не хочет платить за ремонт, аренду и многое другое. Здесь есть общая выгода для обеих сторон.

А как насчет AI-агента, цель которого — минимизировать риск повреждения любых активов, застрахованных по действующим полисам? Этот целеустремленный AI начинает работать, чтобы связаться со страхователями о предстоящем не представляющем угрозы жизни, но разрушительном для имущества шторме. На основе близости, прогнозов и других факторов он дает рекомендации, такие как «Поставьте машину в гараж» или «Есть укрытое парковочное место в трех минутах езды от вашего текущего местоположения; вот маршрут». Мы думаем, вы согласитесь, что большинство страхователей оценят такой подход, если он в их интересах (мы бы оценили). Как оказалось, одна страховая компания, с которой мы работали, использовала этот подход для повышения вовлеченности клиентов и обнаружила, что 52% страхователей пытались предпринять действия по таким оповещениям. И из тех, кто смог предпринять значительные действия, только 6,1% подали иск!

Другие примеры: выявление аномалий, автоматическая оценка убытков, проверка соответствия требованиям, управление утечками по убыткам и совокупным коэффициентом, прогнозирование стоимости рисков, моделирование последствий отказов, персонализированные полисы, цифровое взаимодействие, динамическое ценообразование и скидки, оптимизация первого уведомления об убытке, выявление мошенничества, идентификация новых рыночных возможностей, параметрическое страхование, персонализированное профилирование рисков, разработка продуктов, перестрахование, и многое другое.

Юридическая сфера

GenAI и агенты становятся для юристов такими же необходимыми, как кофе и учебники BARBRI,² формируя будущее этой профессии, один запрос за другим. Почему возможностей для изменений здесь больше, чем в большинстве профессий? Goldman Sachs отметил, что, по их мнению, 44% повседневных юридических задач можно оптимизировать или ускорить с помощью AI — по сравнению со средним показателем по отраслям в 25%. Одной из причин такого вывода является то, что задачи, которые хорошо выполняет AI, совпадают с теми, которые выполняют юристы. Подумайте об этом: задачи, такие как анализ документов, написание, составление аргументов и контрактов, извлечение информации, исследования, проверка и обобщение, идеально подходят для AI.

Представьте себе горы судебных дел и доказательств, необходимых для построения доказательной базы по делу. Как просеять все эти данные? Как один человек может просмотреть 260 000 документов, чтобы полностью понять определенную точку зрения или как детали связаны друг с другом? Это не поиск иголки в стоге сена; это поиск иголки в стоге иголок.

² BARBRI — известная в США компания, которая специализируется на подготовке к экзаменам на адвоката. Она предлагает курсы и материалы для подготовки к экзамену (bar exam), который необходимо сдать, чтобы получить лицензию на юридическую практику в большинстве штатов США. BARBRI предоставляет учебные материалы, лекции, практические тесты и другие ресурсы, чтобы помочь студентам успешно сдать экзамен.

Ситуация станет еще более сложной, потому что и понятие доказательства тоже эволюционировало. Например, в Италии доказательства из WhatsApp (второго по популярности приложения в мире) используются для развода почти половины итальянских изменщиков! Джан Этторе Гассани из Итальянской ассоциации брачно-семейных юристов отмечает: «Мы видели, как изменщики используют эту услугу, чтобы поддерживать три или четыре отношения одновременно, это как динамит». Мы не уверены, что происходит в Италии; однако очевидно, что платформы общения (такие как WhatsApp, Facebook и Instagram) и способы нашего общения (некоторые суды постановили, что отметка «нравится» может юридически обязывать контракты в текстовых сообщениях) фундаментально изменяют объем и определение доказательств.

Теперь представьте, что вы хотите проанализировать все эти документы и приложения, чтобы выяснить, какие люди и компании связаны друг с другом в контексте кредитных соглашений. Вы могли бы использовать агента, который сначала проходит через все ваши документы и использует возможности разрешения именованных сущностей, предоставляемые его поддерживающей LLM, чтобы захватить и связать список людей, мест и вещей. Рабочий процесс затем выводит эти связи в формате markdown,³ а вызов инструмента преобразует markdown в интеллект-карту, которая представляет все стороны в этих документах и как они связаны друг с другом. Внезапно у вас появляется визуализация людей и компаний, которые работают вместе. Это отличный подход для оценки контрагентов (также используется в страховании, банковском деле и т.д.).

Один из наших любимых примеров использования для юристов (как ни странно, он горизонтальный)— возможность «общаться с документом». Откройте документ — например, судебное решение или недавно принятый закон — и взаимодействуйте с этим документом на естественном языке. Например, если вы (или агентство по подбору персонала, с которым вы работаете) используете AI для найма в Нью-Йорке, вам нужно знать о Законе 144, который регулирует использование инструментов автоматизированного принятия решений при найме (AEDTs). Мы использовали Ollama и LLM granite3-dense:2b, чтобы обобщить (из корпуса документов) аспекты, которые нам нужно знать об этом законе. Ответ LLM:

Текст, который вы предоставили, представляет собой набор правил и рекомендаций, касающихся использования инструментов автоматизированного принятия решений при найме (AEDTs) в Нью-Йорке. Вот краткое резюме:

1. Требования к аудиту на предмет смещения: Работодатели и агентства по найму должны проводить аудит на предмет смещения своих AEDT каждые два года, используя независимого аудитора. Аудит должен включать расчеты для всех категорий, кроме тех, которые представляют менее 2% данных.
2. Ответственность за аудит на предмет смещения: Работодатели и агентства по найму несут ответственность за то, чтобы не использовать AEDT без аудита на предмет смещения. Поставщики не могут проводить собственные аудиты на предмет смещения; они должны нанимать независимых аудиторов.
3. Требования к уведомлению: Работодатели и агентства по найму должны уведомлять сотрудников и кандидатов на работу в Нью-Йорке о использовании AEDT и о качествах или характеристиках работы, которые он будет оценивать. Они должны предоставить это уведомление за 10 рабочих дней до использования AEDT.
4. Независимые аудиторы: Нет списка одобренных независимых аудиторов. Они должны быть беспристрастными и не иметь финансовой заинтересованности в работодателе или поставщике.
5. Жалобы: Если кто-то подозревает нарушение, он может позвонить по телефону 311 или подать жалобу на сайте DCWP. Жалобы на дискриминацию следует направлять в Комиссию по правам человека Нью-Йорка.

³ Markdown — это легкий язык разметки, используемый для форматирования текста. Он позволяет добавлять элементы форматирования, такие как заголовки, списки, ссылки и изображения, с помощью простых текстовых символов. Например, чтобы сделать текст жирным, его нужно заключить в двойные звездочки (`***текст**`), а для создания списка используются звездочки или дефисы. Markdown широко используется для написания документации, блогов и сообщений в чатах, так как он легко читается и пишется.

6. Исключения: Некоторые категории могут быть исключены из обязательных расчетов, если они представляют менее 2% данных.

Посмотрите, как LLM быстро обобщила корпус документов, который мы ей предоставили? Теперь вы можете адаптировать эту информацию и мгновенно подготовить ее для клиента, чтобы сформировать общение с целевой аудиторией. Например, это может быть рассылка от Бюро лучших бизнес-практик в приветственном наборе для новых магазинов, открывающихся в городе, отправленная целевой аудитории, которая не может понять юридический жаргон, или, возможно, сообщения для людей, которые не являются носителями английского языка.

Как и в других профессиях, юристы не исчезнут из-за AI. Наоборот, мы считаем, что спрос на юридические консультации и услуги только возрастет, учитывая все возрастающую сложность мира. GenAI сам по себе создал множество юридических вопросов, связанных с честным использованием авторских прав. Тем не менее, мы склонны согласиться с тем, к чему стремился Goldman в своем отчете: использование AI для выполнения первоначального этапа «тяжелой работы» значительно снижает стресс и монотонность (сдвиг влево), позволяя юристам минимизировать время, затрачиваемое на предварительные, низкоуровневые проверки, и заниматься более интеллектуальной работой (сдвиг вправо). Конечно, в этой отрасли (как и в других, таких как здравоохранение) есть серьезные подводные камни, связанные с AI, и мы рассмотрим некоторые из них в следующей главе.

Другие примеры использования AI в юриспруденции включают прогнозирование арбитража, автоматическую проверку должной осмотрительности, поддержку альтернативного разрешения споров, переписку, составление и изменение контрактов и юридических документов, трансформацию обслуживания клиентов, управление жизненным циклом контрактов, проверку документов, электронное открытие дел, повышение операционной эффективности, облегчение доступа к знаниям, преодоление языковых или профессиональных барьеров, юридические исследования и анализ документов, управление удержанием и сохранением электронной информации, поддержка переговоров, прогнозирование потенциальных исходов судебных разбирательств или обязательных арбитражных решений, упрощение приема дел через веб-интерфейс и др.

Производство и промышленность

Промышленность — одна из тех отраслей, где смелые лидеры стремятся к изменениям. От цепочек поставок, которые иногда кажутся сложнее, чем ваша школьная любовь, до тяжелого оборудования. Не забывайте также о производственных линиях, людях и процессах, управлении объектами и многом другом. Мы уверены, что вы не смогли бы пройти по производственному предприятию, не заметив дюжину проблем, которые можно было бы решить или облегчить с помощью GenAI.

Отличный пример того, насколько широко распространен GenAI в промышленности, — составление финансовых отчетов. Основные средства (PP&E) являются частью обязательных ежегодных финансовых отчетов компании, представляющих ее повседневную деятельность. Используйте AI для создания первого черновика на основе прошлогоднего отчета, динамически обновленного данными, метриками и журнальными записями за текущий год. Очевидно, что финансовая отчетность требует контроля, поэтому участие человека критически важно, но это отличное начало. Более того, со временем (используя такие методы, как InstructLab, о которых мы расскажем в главе 8), вы можете обновлять свою модель с учетом изменений в правилах раскрытия информации в соответствии с общепринятыми принципами бухгалтерского учета (GAAP), которые, возможно, изменились с прошлого года. Мы понимаем, это звучит как пример из бухгалтерии... потому что это он и есть. Мы включили его сюда, потому что это отличный пример бухгалтерии в производстве. Помните, что в бизнесе можно комбинировать как горизонтальные, так и вертикальные примеры использования.

Аддитивное производство (Additive Manufacturing, AM) — это термин, описывающий процесс создания изделий путем послойного добавления материалов. Эти слои могут быть из пластика, порошков, металла, бетона — и однажды даже из человеческих тканей! AM уже применяется в авиации, стоматологии (восстановление зубов), медицине (импланты), автомобилестроении и даже моде. Как правило, процессу AM предшествует использование программного обеспечения для моделирования (например, систем автоматизированного проектирования CAD), которое передает эскиз в AM-систему, где начинается послойное наращивание изделия.

Ранее в этой книге говорилось о том, как безъязыковые (nonverbal) или предметно-ориентированные языки программирования трансформируются в LLM. Пример: Georgia Tech создали polyBERT — LLM, «мыслящую» о химической структуре полимеров как о языке. Это похоже на инициативу L'Oréal, описанную во 2-й главе, но здесь язык сосредоточен не на косметике, а на полимерах. Эта LLM может делать выводы о формах и свойствах и предсказывать поведение веществ.

Использование LLM для проектирования CAD-документов или поиска новых связующих веществ — отличный кейс для производства. Сомневаетесь? Спросите Брайсона Дешамбо, который использовал инновационные 3D-печатные клюшки и одержал историческую победу на US Open по гольфу в 2024 году. Традиционно гольф-клюшки высшего класса куют (раскаленный металл прессуется в форму), но аддитивный подход клюшек Брайсона позволил его команде постоянно улучшать дизайн с недоступной ранее скоростью. Представьте себе LLM, созданную для гольфа, только без нецензурной лексики.

AI также может быть применен для учета эксплуатационных факторов (использование, обслуживание, стоимость материалов) с целью составления оптимального производственного графика, минимизирующего простои и максимизирующего доступность оборудования.

В производстве, похоже, единственный предел — это найти проблему, которую нужно решить. Вот еще примеры таких задач: обнаружение загрязнений при выпекании, соблюдение нормативных требований, анализ цвета, дозировка химических веществ, энергоэффективность, производительность оборудования, разработка и открытие новых материалов, повышение точности выявления дефектов, предиктивное техническое обслуживание, характеристика полимеров, контроль правильности экипировки (и других аспектов безопасности сотрудников), выявление и прогнозирование дефектов, достижение целей устойчивого развития, снижение незапланированных простоев и многое другое.

Фармацевтика

Как и любая другая отрасль, фарма сейчас переживает свой «AI-glow-up» — только в этом случае речь идет не просто о сэкономленном времени на учет пяти пропущенных статусных звонков в бэклоге, а о спасении жизней.

Подумайте о том, какие выводы можно извлечь из огромного объема данных в этой отрасли. Возможности поистине безграничны, особенно если учесть, что компании обычно не публикуют свои неудачи при разработке лекарств или во время клинических испытаний. Чему бы мы могли научиться, если бы AI получил доступ к исследованиям о препаратах, которые так и не вышли на рынок? Производство в сфере наук о жизни генерирует большие объемы данных, которые, как правило, разбросаны по внутренним и внешним системам, лишенным совместимости и согласованности — какие инсайты могли бы быть извлечены для улучшения качества жизни пациентов или улучшения исходов лечения?

А теперь задумайтесь о сценариях применения AI, связанных с ускорением («сдвигом влево») сроков разработки лекарств, если учесть, что в среднем новые препараты выходят на рынок через 10–14 лет после начала работ, а совокупные затраты составляют около 2,6 миллиарда долларов. Более того, только 8–10% препаратов, доходящих до клинических испытаний, в итоге проходят все стадии и доходят до рынка. Подумайте об этом: в этой отрасли, если вы терпите неудачу «всего лишь» в 80% случаев (вдвое увеличив свой шанс на успех до 20%), вы уже герой сдвига-влево!

На сегодняшний день большинство клинических испытаний сталкиваются с неэффективностью площадок по ряду причин. Во-первых, увеличение количества участников отрасли и сужение популяций требует все более специфических критериев отбора участников в исследования. Это означает, что гонка «кто первым найдет — тот и вылечит» стала конкурентной как никогда. На самом деле набор участников на исследование средней фазы может занимать до 10 месяцев — AI может сократить это время вдвое. Выбор площадки (и соответствующие затраты) — еще один фактор.

Но, пожалуй, самой большой проблемой испытаний остается привлечение и удержание пациентов. Крупные фармкомпании испытывают трудности с идентификацией, набором и удержанием разнородной аудитории пациентов. Эта фаза может занимать до 30% всего времени разработки препарата и способна задержать исследования на месяцы. И что еще важнее — как только пациент включен в исследование, появляется необходимость ускоренного внедрения новых методов и подходов для улучшения его опыта и удержания (еще один существенный «узкий участок», вносящий

вклад в задержки). Теперь подумайте горизонтально: как AI может быть использован для поддержки аналитически-ориентированной стратегии набора, персонализированного подхода к набору и удержанию пациентов, динамического и прогностического мониторинга исследовательских площадок.

Что ждет нас в будущем, вызывает огромный интерес, особенно с учетом того, как все большее значение приобретает квантовые вычисления, способные ускорить разработку лекарств. Например, примерно 80–90% белков в организме по-прежнему недоступны для современных лекарств, поскольку экспериментальные препараты не могут «пристыковаться» к ним (то есть не образуют необходимую силу связывания), чтобы повлиять на нужный биологический путь. Возьмем, к примеру, всем известный пенициллин: моделирование его структуры потребовало бы классического компьютера с числом транзисторов, превышающим количество атомов во всей наблюдаемой Вселенной — физически невозможная задача. Но это уже лежит в области квантовой реальности. Какие новые соединения и препараты сможет открыть мир?

Вот еще несколько примеров применения GenAI и агентных систем в фармацевтической отрасли, способных вдохновить на более глубокое осмысление ее трансформационного потенциала: отчетность по нежелательным побочным явлениям, обнаружение загрязнений в производственных партиях, интеграция по непрерывной цепочке ухода, оптимизация логистики «холодовой цепи», разработка новых химических соединений, динамическое управление запасами, расширенное обучение пациентов с адаптацией сложности инструкций, повышение разнообразия участников клинических исследований, оптимизация лекарственных формул, оптимизация производственных процессов, прогнозирование лекарственных взаимодействий, репрофилирование (репозиционирование) существующих препаратов, симуляция обеспечения качества, и многое другое.

Безграничные возможности: другие отрасли, где GenAI проявляет себя

Когда мы писали эту главу, мы уже знали, что издатель сократит ее. Не потому что в ней нет полезной информации, а потому что на одну только эту тему можно было бы написать целую книгу. Существует множество отраслей, которые мы не охватили в этом разделе (возможно, это идея для следующей книги, которую мы пообещали себе не писать). Тем не менее, мы не могли закончить этот раздел без короткого обзора того, как GenAI и агентные технологии производят фурор в некоторых других индустриях:

Автомобилестроение

Системы помощи водителю (ADAS), взаимодействие с клиентом после покупки, оптимизация спецификаций, персонализированные функции «по воздуху», подключенные ассистенты водителя, непрерывное тестирование и верификация, анализ поведения потребителей, исследование инженерных решений, привлечение клиентов, удаленное обслуживание и диагностика, аналитика по безопасности, оценка стоимости аварийных авто, умные автомобили, управление цепочками поставок, все сценарии из производства, и так далее.

Банковское дело и управление капиталом

Ускоренная обработка заявок на кредит, борьба с отмыванием денег, соответствие требованиям, тестирование пороговых значений дневных депозитов, анализ клиентов и рекомендации по кросс-продаже, давление на комиссионные, расчет рейтингов, проверка клиентов, отчетность по запросу, выявление и предотвращение мошенничества, инвестиционные советы, максимизация спреда по депозитам, контроль рисков, открытие счетов и проверка рисков, анализ расходов, налоговая эффективность, анализ настроений, анализ законодательства, технологически поддерживаемый аудит, торговые риски, корректировки оценок, и так далее.

Розничная торговля

Автоматизированное управление цепочками поставок, управление кампаниями, цифровая мода (да, украшения Gucci для вашего аватара, нет, мы туда не пойдём... никогда), подбор размеров и примерка, анализ трафика в магазинах, гиперперсонализированный маркетинг, виртуальный консьерж в магазине, управление запасами, подбор по вкусу и параметрам клиента, программы лояльности, оптимизация цен, логистика покупок, предиктивное обслуживание, рыночная аналитика в реальном времени, платформа управления магазином, дизайн стиля, усиленные продавцы, виртуальные шоурумы, и так далее.

Государственные услуги

Обслуживание граждан, оборона и национальная безопасность, экологическая аналитика, борьба с мошенничеством, судебная система, инклюзивность жизненно важных сервисов, обслуживание инфраструктуры, общественная безопасность, разработка и анализ политик, суммирование законодательных актов и слушаний, эксплуатация объектов недвижимости, составление технических заданий и контрактов, слушания, оптимизация трафика, выявление налоговых разрывов, управление отходами и переработка, учет вакцинации и оценка общественного мнения, виртуальные госслужащие, и так далее.

Строительные блоки ИИ

К чему все это ведет? Вспомните главу 3, где мы представили нашу формулу успеха ИИ (Успех ИИ = Модели + Данные + Управление + Сценарии использования). Теперь, дойдя до этой части книги, мы продвинулись дальше. Мы эволюционировали от «формулы» к «каркасу» успешного ИИ: строительным блокам ИИ. И если Кривая создания ценности ИИ была для вас вдохновением, то строительные блоки ИИ (рис. 4.3) станут вашим фундаментом. Бизнес-ценность должна начинаться со сценариев использования, потому что лучший способ реализовать возможности ИИ — это говорить о том, что он значит для вашего бизнеса. Помните: согласование бизнес-стратегии и технологической стратегии — это ваша путеводная звезда.

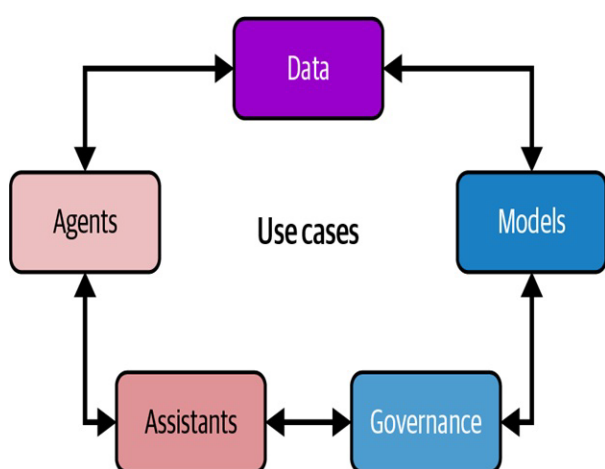


Рис. 4.3. Строительные блоки ИИ

Чтобы не отставать от технологий и быстрее переходить от экспериментов к внедрению, мы хотим, чтобы вы всегда думали об этих строительных блоках ИИ. На этом этапе книги мы посчитали уместным объединить все вместе, чтобы завершить главу:

Сценарии использования должны отвечать на вопрос: каких целей я хочу достичь с помощью ИИ в своем бизнесе? Это и есть фокус данной главы – сценарии использования с бизнес-ценностью, а не технические проекты ради интереса.

Данные. Сильная основа данных критически важна для того, чтобы данные работали на ваш бизнес. И речь не только о данных, на которых обучается модель; это также организация, понимание и управление данными. Без этого вы толком не знаете, какие данные у вас есть (если вообще можете их найти), можно ли их безопасно использовать, какие у них правила хранения и удаления, и многое другое. Ваши данные подготавливаются к ИИ с помощью информационной архитектуры (IA), которую мы обсуждали в главе 2 и подробнее рассмотрим с точки зрения ее преимуществ в главе 8. Вашему ИИ нужна IA.

Модели. Правильные модели — это необходимость. Этот строительный блок включает в себя такие аспекты, как экономическая эффективность, возможность развертывания, доверие, прозрачность, открытость и производительность. Появился новый класс LLM — так называемые малые языковые модели (SLM). Вы уже немного познакомились с их эффективностью, когда мы говорили о помощниках для программирования с поддержкой ИИ. Появляются новые архитектуры, такие как Mixture of Experts (MOE), которые используют разработчики моделей, например IBM, DeepSeek и Mistral. Ваша компания не будет определяться одной моделью. Это тема главы 7. Одна модель не сможет все.

Управление. Вам нужно полноценное управление ИИ от начала до конца, чтобы вы могли с уверенностью внедрять его по всему предприятию. Это также необходимо из-за все большего

количества нормативных требований, связанных с использованием ИИ. Это тема главы 5. Доверие станет лицензией на работу, потому что ИИ, которому доверяют, — это ИИ, который будут использовать.

Ассистенты дают еще больше продуктивности. Люди обеспечивают возможности, а ассистенты и другие формы цифровой рабочей силы — масштабируемость.

Агенты поднимут эту масштабируемость на новую высоту продуктивности благодаря своей большей автономности и способности учиться, запоминать и корректировать действия. Хотя некоторые сценарии использования были приведены в этой главе, подробнее об агентах мы поговорим в главе 7. Агенты откроют следующую волну роста продуктивности для бизнеса.