ГЛАВА 1

Agentic mesh: основы

За несколько месяцев до начала работы над книгой появились массовые сообщения о десятках миллиардов долларов вложений в интеллектуальных агентов, включая инфраструктуру их поддержки, инструментарии их создания, предназначенные для них LLM-модели, обучение инженеров и других специалистов по работе с агентами, и, конечно, огромное количество роликов на YouTube, расхваливающих эти невероятные новые возможности.

Тем не менее агенты в той или иной форме существуют уже давно. Собственно, если заглянуть достаточно далеко в прошлое, ранние концепции агентов можно обнаружить в трудах Алана Тьюринга, который представил машинный интеллект и тест Тьюринга для оценки способности машины демонстрировать интеллект, похожий на человеческий. Эта фундаментальная работа заложила основу для того, что мы сегодня называем интеллектуальными агентами.

Несколько десятилетий спустя начали появляться первые реализации очень примитивных агентов. В 1980-х стали доступными первые экспертные системы, имитирующие принятие решений человеком. Затем пришел черед виртуальных помощников (таких, как Siri от Apple и другие), которые стабильно развивались. Но только в 2017 году область по-настоящему преобразилась, когда исследователи Google опубликовали эпохальную статью «Attention Is All You Need»[1] с описанием архитектуры, совершившей революцию в ИИ. Этот прорыв заложил фундамент для современных продвинутых LLM, благодаря которым агенты получили возможность обрабатывать, генерировать и взаимодействовать так, как это делают люди.

Тем не менее современные ИИ-агенты представляют качественный прорыв, а не инкрементное улучшение. В отличие от своих предшественников, агенты на базе LLM способны рассуждать, понимать контекст диалогов, генерировать творческие решения и адаптироваться к новым ситуациям, не имеющим заранее запрограммированных ответов.

Мы начали разрабатывать концепции и архитектуры, описанные в книге, незадолго до появления агентов на базе LLM. Наша работа строилась на более ранних экосистемах — таких, как сетки сервисов (service mesh) для API и сетки

[1] «Внимание — все, что вам нужно».

данных (data mesh) для продуктов данных. Эта книга расширяет родословную агентов до того, что мы называем *агентной сеткой* (agentic mesh). Экосистема в простейшем виде представляет собой набор взаимосвязанных частей, зависящих друг от друга. В технологии экосистемы появляются там, где компоненты — будь то сервисы, данные или агенты — спроектированы для совместной работы.

Впрочем, термин «агент» имеет разный смысл для разных людей, поэтому мы заранее договоримся о следующих определениях, чтобы избежать возможных разногласий.

Агентный

Способный принимать независимые решения — как правило, для выполнения цели.

Агентный ИИ

Осуществляет сложные рассуждения и итеративное планирование с использованием LLM для автономного решения многошаговых задач.

Агенты

Используют агентный ИИ для независимого планирования и выполнения задач.

Объединив их, мы получаем простое определение агента.

> *Агент — это программа, использующая LLM и способная принимать самостоятельные решения, осуществлять итеративное планирование и выполнять задачи для достижения сложных целей.*

А когда речь заходит о «сетке» агентов («агентной сетке»), имеется в виду следующее.

> *Агентная сетка — взаимосвязанная экосистема, которая помогает агентам находить друг друга, сотрудничать, взаимодействовать и осуществлять транзакции.*

В агентной сетке основными компонентами являются агенты. Они проектируются с учетом факторов управления, взаимодействия и доверия, чтобы они могли сотрудничать, взаимодействовать и даже выполнять операции в более широкой экосистеме агентов. Однако между потребностями отдельного агента и большой экосистемы имеется принципиальное отличие. Экосистемы существуют для того, чтобы сделать возможным сотрудничество в масштабе, когда тысячи агентов, каждый из которых работает независимо, могут планировать, выполнять задачи и выдавать согласованные результаты. Для решения таких проблем были созданы агентные сетки.

Разобравшись с ситуацией, посмотрим, как мы к ней пришли.

Появление LLM

В 2023 году компания OpenAI выпустила LLM GPT-3.5, которая — возможно, впервые — позволила ИИ вести разговор в относительно человеческой манере. С того времени темп только нарастал. Возможности LLM стремительно расширялись, а затраты на них снижались еще быстрее. Многие когда-то небольшие отраслевые компании — OpenAI, Anthropic и другие — а также технологические гиганты (в частности, Google и Microsoft) запускали такие сервисы, как ChatGPT и Claude, которых ожидали беспрецедентный рост и признание.

Это быстрое развитие радикально изменило то, как люди и организации взаимодействуют с ИИ. Инструменты, некогда требовавшие значительного технического опыта, в наши дни доступны для широкой публики, и миллионы людей могут внедрять нетривиальные возможности ИИ в свою частную и профессиональную жизнь. Благодаря своей способности генерировать текст, анализировать данные и предоставлять поддержку клиентам LLM наделяют повседневные приложения человекоподобным разумом.

Но что это значит? В контексте LLM «интеллект, подобный человеческому» означает способность модели понимать язык, делать заключения относительно контекста и генерировать ответы, близкие к тому, как люди мыслят и общаются. Например, когда клиент спрашивает у чат-бота авиакомпании «Мой рейс задерживается, и я не успеваю на пересадку — что мне делать?», система не просто ищет ключевые слова — она интерпретирует намерения, оценивает ситуацию и отвечает решением (например, переносит бронь или направляет к ближайшей службе поддержки). Такое сочетание понимания, контекстных рассуждений и адаптивного решения задач воспроизводит то, как подобный запрос обрабатывался бы человеком; именно поэтому мы описываем этот интеллект как человекоподобный.

В результате этих невероятных достижений многие специалисты прогнозируют значительное повышение производительности и возможностей для роста. Лила Ибрагим (Lila Ibrahim), директор по эксплуатации Google DeepMind, пишет: «У ИИ существует огромный потенциал для преобразования нашего мира к лучшему. От ранней диагностики болезней и разработки новых лекарств до решения важных экологических проблем за счет открытия новых экологичных материалов, ИИ уже способствует прогрессу в решении некоторых самых сложных проблем нашего общества».

Ибрагим подчеркивает ощутимое влияние, которое ИИ уже оказывает на здравоохранение, научные исследования и экологию. Однако эффект ИИ охватывает все отрасли. Организация экономического сотрудничества и развития (OECD) говорит, что «ИИ уже производит революцию во многих критически важных бизнес-функциях в разных секторах, таких как рекомендации контента, онлайновые продажи и поддержка клиентов... Его способность к анализу данных, автоматизации процессов и улучшенному принятию решений обещает экономический рост и социальный прогресс».

На основании этой информации консультационная компания McKinsey утверждает, что «генеративный ИИ может запустить следующую волну роста производительности... Влияние GenAI на производительность может добавить в глобальную экономику вклад, измеряемый триллионами долларов». По оценкам последних исследований McKinsey, GenAI «может ежегодно добавлять в экономику эквивалент от 2,6 до 4,4 триллиона долларов». И еще McKinsey утверждает, что «генеративный ИИ оказывает значительное влияние во всех экономических секторах. Банковское дело, высокие технологии и биология принадлежат к числу отраслей, которые могут извлечь наибольшую выгоду от GenAI в процентном отношении к доходам... Например, в банковской отрасли при полной реализации практических сценариев технология может ежегодно создавать ценность, эквивалентную дополнительным 200–340 миллиардам долларов. В области розничной торговли и производства потребительских потенциальный эффект также весьма значительный — по некоторым оценкам, он составит от 400 до 600 миллиардов в год».

Вполне естественно, что эти перспективы не остаются незамеченными. Вложения в инфраструктуру, относящуюся к ИИ (дата-центры, фундаментальные модели), стремительно растут. По оценкам JPMorgan, «в настоящее время в США строится более 3800 МВт мощности, что на 70 % больше, чем в предыдущем году, и еще 7000 МВт находятся на различных стадиях подготовительных работ». В другой статье Reuters утверждает: «Вложения в дата-центры, которые будут предоставлять вычислительные мощности для ИИ, резко выросли с момента запуска ChatGPT компанией OpenAI в 2022 году, так как компании в разных секторах стали все чаще перемещать свои операции на облачные платформы и интегрировать ИИ в свои предложения».

Эта волна инфраструктурного роста предполагает, что существует намного более широкая возможность интеграции ИИ практически в любые бизнес-процессы. Организации все чаще перемещают свои рабочие нагрузки в облако и применяют решения на базе ИИ для рационализации процессов, оптимизации ресурсов и улучшения опыта клиентов. Как бы то ни было, масштаб этих вложений демонстрирует, что компании понимают фундаментальные последствия развития ИИ и стремятся не упустить возможность воспользоваться его потенциалом.

Мощь ИИ изменяет наш мир с беспрецедентной скоростью. ChatGPT достигает 800 миллионов еженедельных активных пользователей всего за 17 месяцев, став самым быстрорастущим продуктом в истории и продемонстрировав темпы внедрения, на фоне которых блекнут даже самые успешные предприятия эпохи интернет-бума. Тем временем глобальные вложения достигли невероятных уровней: шесть крупнейших технологических компаний США повышали свои капитальные затраты, относящиеся к ИИ, приблизительно на 63 % в год, до 212 миллиардов долларов в 2024 году, тогда как объявления о вакансиях, связанных с ИИ, к 2018 году выросли на 448 %.

Тем не менее при всех значительных достижениях ИИ и LLM сейчас мы наблюдаем лишь начало намного более масштабного преобразования. В отличие от

предыдущих технологических волн, на становление которых ушли десятилетия, развитие ИИ происходит в невиданном темпе, сопровождаясь повышением производительности моделей. В то же время затраты на инференс упали на 99,7 % всего за два года. Технология продолжает развиваться с головокружительной скоростью, и мир входит в эпоху, когда ИИ не просто помогает в завершении задач — он автономно принимает на себя обязанности, решает задачи и создает возможности, которые на фундаментальном уровне переопределяют сам процесс выполнения работы.

Эпоха агентов

Мы предвидим эволюцию ИИ с переходом от реактивных инструментов к проактивным, автономным агентам — сущностям, которые не только реагируют на промпты, но и самостоятельно планируют многошаговые рабочие потоки, выполняют их и принимают решения в пределах определенных параметров. Начали появляться первые примеры такого рода: ИИ-агенты, согласующие встречи между несколькими участниками с разными календарями, проводящие исследования с просмотром сотен веб-сайтов или осуществляющие поддержку клиентов от исходного контакта до решения проблемы. Они представляют фундаментальный переход от ИИ, отвечающего на вопросы, к ИИ, достигающему целей.

Эти агенты, основанные на тех же достижениях LLM и питаемые беспрецедентными вложениями в инфраструктуру ИИ, сделают обсуждавшиеся ранее коммерческие выгоды не только возможными, но и повсеместными. Они размывают границы между автоматизацией и независимым действием, приводя к будущему, в котором машины сотрудничают с людьми для раскрытия незадействованного потенциала.

Вскоре мы приведем определение и разберемся, что же понимается под агентами, но пока посмотрим, какие последствия предполагаются от их внедрения. Марк Бениофф (Marc Benioff), исполнительный директор крупной технологической фирмы Salesforce, говорит: «Мы входим в новую эру автономных ИИ-агентов, которые самостоятельно выполняют действия и дополняют работу человека. Это не просто технологическое развитие. Это революция, которая на фундаментальном уровне переопределит то, как люди работают, живут и взаимодействуют друг с другом».

Ожидается, что агенты не только улучшат производительность, но и кардинально изменят сам рабочий процесс. Бениофф указывает, что «сегодня мы уже привыкли к “прогностическому ИИ”, который анализирует данные для предоставления рекомендаций, прогнозов и аналитических выводов, и к “генеративному ИИ”, который учится на данных и использует закономерности для органичной генерации текста, изображений, музыки и кода. Агенты — программные компоненты, которые выходят далеко за эти границы. Они могут выполнять задачи самостоятельно, принимать решения и даже вступать в переговоры с другими агентами по нашему поручению».

Возможно, впервые в истории роли поменялись: до сих пор люди определяли развитие технологий. Мы — люди — держали руль и направляли технологию в нужном направлении. Теперь агенты могут стать технологией, которая будет сама решать, в каком направлении идти (желательно с защитными мерами, но об этом далее в книге). Бениофф продолжает: «Технология не просто предлагает людям инструменты для выполнения работы. Она предоставляет разумную, масштабируемую цифровую рабочую силу, способную выполнять задачи автономно. Вместо того чтобы ожидать указаний от человека, агенты могут анализировать информацию, принимать решения и совершать действия самостоятельно, адаптируясь и обучаясь на ходу».

Всемирный экономический форум — экономический аналитический центр — также прогнозирует, что агенты станут движущей силой крупных экономических изменений: «ИИ-агенты становятся все более совершенными, что имеет значительные последствия для принятия решений, подотчетности и контроля. <...> К преимуществам ИИ-агентов относятся прирост производительности, специализированная поддержка и улучшенная эффективность в таких секторах, как здравоохранение, клиентский сервис и образование».

Несомненно одно: агенты войдут в нашу жизнь. Скоро. И это будет масштабным явлением.

Определение понятия «агент»

Но что же такое «агент»? Рассмотрим несколько определений от ведущих компаний и лидеров мнений.

- «ИИ-агент — система или программа, способная автономно выполнять задачи по поручению пользователей или других систем, планируя собственный рабочий поток и используя доступные инструменты». Источник: IBM.
- «ИИ-агенты — разновидность систем на базе искусственного интеллекта, способных понимать запросы клиентов и реагировать на них без вмешательства человека». Источник: Salesforce.
- «ИИ-агент — система, использующая LLM для определения программной логики приложения». Источник: LangChain.
- «ИИ-агенты — системы, способные к логическим рассуждениям, которые могут понимать контекст, планировать рабочие потоки, связываться с существующими инструментами и данными и выполнять действия для достижения определенной цели». Источник: Deloitte.

К сожалению, многие из этих определений, на наш взгляд, склоняются к так называемому agent-washing (маркетинговое использование слова «агент»). Этим термином описывается нарастающая тенденция компаний-разработчиков называть старые инструменты — чат-боты, макросы, скрипты RPA (Robotic Process Automation) — «агентами», несмотря на отсутствие истинной

автономности. Хотя настоящие агенты могут рассуждать, планировать, действовать и адаптироваться динамически, многие так называемые агентные решения представляют собой обычных ИИ-помощников в новой упаковке. Garther и другие предупреждают, что только небольшая часть современных приложений являются действительно агентными: системами, в которых LLM управляют их собственными процессами, принимают решения и сотрудничают самостоятельно. В отличие от реактивных инструментов, настоящие агенты похожи на людей своей способностью хранить контекст, управлять приоритетами и работать в командах или экосистемах.

Это отличие имеет серьезные архитектурные последствия — все они будут более подробно рассмотрены в следующих главах. Реальные агенты являются headless (не имеют пользовательского интерфейса), распределены, управляются событиями и безопасны — что требует контейнеризации, оркестрации, наблюдаемости и фреймворков идентификации с нулевым доверием. Мы предложим архитектурную модель для агентов, объединяющую микросервисы с механизмами логических рассуждений, которые делают возможным масштабируемое, безопасное и управляемое целями поведение. Компании должны требовать прозрачности и избегать шумихи. Без должной дисциплины они рискуют вложиться в старые инструменты с новым названием, не способные обеспечить ни автономности, ни окупаемости. Чтобы агенты оправдали возложенные на них надежды, необходимо отбросить броские ярлыки и обратиться к архитектурам, которые действительно работают.

Как видите, здесь прослеживаются некоторые общие аспекты. Давайте объединим их с несколькими идеями, чтобы предложить целостное определение агентов.

- Агенты работают на базе моделей LLM, которые обеспечивают их способность планировать и выполнять задачи.
- Агенты имеют последовательный набор характеристик, определяющих их назначение, поведение, масштаб и подотчетность.
- Агенты обладают автономностью и могут действовать самостоятельно в границах их назначения и ограничений, установленных их владельцем.
- Агенты могут пользоваться инструментами и сотрудничать с другими агентами для решения своих задач.

Чтобы избежать путаницы со старым использованием термина «агент», поясним, что мы имеем в виду агентов на базе ИИ (хотя, наверное, это понятно без объяснений).

Агенты сегодня

Как же агенты создаются сегодня? Сейчас используются только примитивные агенты, а большинство организаций продолжают применять *рабочие потоки (*workflows) и полагаются на методичную, предварительно определяемую

оркестрацию LLM и других инструментов (рис. 1.1). Когда у вас имеется рабочий поток, то разработчик или инженер заранее решает, какие инструменты будут использоваться системой на каждом шаге, и четко определяет последовательность действий.

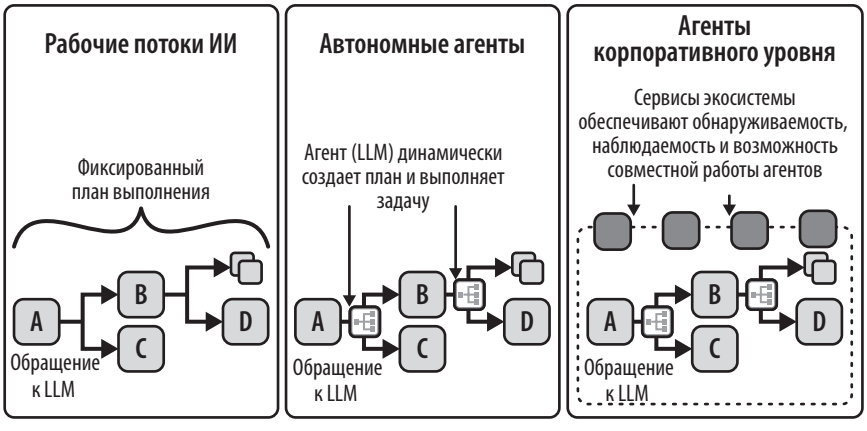


Рис. 1.1. Рабочие потоки ИИ, автономные агенты и агенты в корпоративных средах

Anthropic, компания-разработчик ИИ/LLM, определяет рабочий поток как «систему, в которой LLM и инструменты оркеструются по заранее определенным кодовым путям». Рабочий поток ИИ представляет собой структурированный пошаговый процесс, который интегрирует различные инструменты, модели и алгоритмы для выполнения задачи или набора задач. В отличие от агентов, которые проектируются для работы с определенной степенью автономности и адаптивности, рабочие потоки полагаются на заранее заданные инструкции и фиксированные точки принятия решений, установленные разработчиками.



Рабочий поток ИИ можно наглядно представить в виде сборочной линии на автозаводе: каждый шаг тщательно проектируется с целью продвижения задачи по фиксированному пути, с минимальной свободой для отклонений или самостоятельного принятия решений.

Даже в этом случае рабочие потоки ИИ обладают достаточно мощными возможностями. Anthropic определяет некоторые стандартные паттерны рабочих потоков.

- *Цепочки промптов* разбивают задачи на последовательные этапы, при этом каждый этап обрабатывает выходные данные предыдущего этапа — например, генерация рекламного текста, его перевод на другой язык и проверка точности.
- *Рабочие потоки с маршрутизацией* классифицируют входные данные и направляют их специализированным процессам — например, запросы клиентов разделяются на запросы возврата средств, технические проблемы и общие вопросы.

- *Рабочие потоки с параллелизацией* разделяют задачи на независимые подзадачи, делая возможной одновременную обработку — например, анализ разделов документа или агрегирование выходных данных нескольких запусков.
- *Рабочие потоки «оценщик — оптимизатор»* производят итеративное уточнение результатов, при котором одна LLM генерирует контент, а другая предоставляет обратную связь по аналогии с редактированием с участием людей.

Но для чего особо выделять рабочие потоки? Потому что они невероятно мощны, и ваши агенты, вероятно, будут их использовать. Однако рабочие потоки отличаются от агентов. Современные рабочие потоки — предшественники агентов.

Агенты, в отличие от рабочих потоков, динамически создают собственный план выполнения задачи: они выбирают инструменты, пути выполнения и управляют выполнением задачи. В отличие от рабочих потоков, агент обладает встроенной способностью определить, как лучше выполнить задачу без заранее заданной статической реализации. Это означает, что агент может на ходу решать, когда выполнять вычисления, обратиться к базе данных или иным образом адаптировать свой план.

Вместо того чтобы следовать фиксированной последовательности шагов, агенты анализируют задачи в реальном времени, генерируют собственные планы и выбирают наиболее подходящие инструменты и шаги выполнения для достижения целей. Эта гибкость позволяет агентам адаптироваться к новым или непредвиденным ситуациям, вследствие чего они особенно хорошо подходят для сложных, открытых задач. Например, агент, которому поручен сбор информации о рыночных трендах, может динамически решить, что ему стоит обратиться к конкретным базам данных, проанализировать недавние новостные статьи и провести анализ тональности публикаций в социальных сетях либо вступить в сотрудничество с другими агентами, адаптируя свой подход при появлении новой информации.

Хотя настоящая автономность агентов все еще находится на ранних стадиях развития, современные разработчики активно объединяют предсказуемость рабочих потоков с гибкостью, характерной для агентов. Такие «примитивные агенты» часто работают в четко определенных границах, используя рабочие потоки как фундамент, который дополняется элементами принятия решений и адаптации. Мы полагаем, что эти ранние реализации закладывают основу для более сложных агентов, демонстрируя практическую ценность объединения статических рабочих потоков с динамически изменяемой функциональностью.

Агенты корпоративного уровня

Итак, хотя агенты создаются уже сегодня, остается ряд насущных вопросов. Готовы ли они занять свое место в технологической среде организации? Окупится ли их эксплуатация? Можно ли отнести их к продуктам корпоративного уровня?

Мы полагаем, что с ростом автономности и сложности агенты должны перейти на корпоративный уровень. Они должны легко интегрироваться в ландшафт

технологий и приложений компаний. Они должны соблюдать правила корпоративных процессов (например, DevSecOps и MLOps), которые обеспечивают дисциплину, необходимую для вывода критически важных приложений, а затем и агентов, в продакшен. Они должны отвечать требованиям всех корпоративных приложений по обнаруживаемости, наблюдаемости, управляемости, безопасности и достоверности.

Задумайтесь: если создаваемые нами агенты не готовы к эксплуатации, если они не обладают возможностями, общими для большинства корпоративных сред, если они не соответствуют корпоративным процессам, то они никогда не перейдут в продакшен. И в этом случае они не смогут приносить пользу. На наш взгляд, это не только большая проблема, но и почти экзистенциальный вызов, который должен быть преодолен для успешного развития агентов.

В будущем мире агентов нужны агенты корпоративного уровня. В наши дни корпорации обычно строят уникальные, ситуативные решения для подключения агентов к продакшен-среде. Такие разработки явно не самодостаточны, что приводит к созданию точечных решений и, как следствие, к неизбежному росту технического долга.

Итак, хотя ситуация кажется очевидной, о ней необходимо еще много сказать. Агентная архитектура корпоративного уровня (рис. 1.2) включает несколько ключевых компонентов.

Эндпоинты

Доступ к агенту осуществляет через хорошо понятную функциональность, которая обычно используется в микросервисах (например, REST).

Базовые возможности

Агенты обладают обнаруживаемостью (discoverability), наблюдаемостью (observability), управляемостью (operability) и достоверностью (trustfulness). Это позволяет легко находить их, отслеживать их работу и заставлять выполнять операции, доверяя им при этом.

Безопасность

Во всех микроагентов встраиваются стандартные корпоративные функции (или микроагенты используют эти функции): взаимный TLS, OAuth2 для управление доступом на основе ролей, интеграция с реестрами идентичностей (identity book of record).

Сотрудничество

Агенты могут сотрудничать с людьми или другими агентами. Они могут идентифицировать агентов-сотрудников, предоставляющих необходимые возможности, общаться на естественном языке, управлять состоянием в продолжительных запросах и взаимодействовать с людьми или другими агентами, когда необходимо получить дополнительную информацию.

Управление задачами/сбор информации

Агенты способны динамически планировать и выполнять задачи. Они могут использовать различные интеллектуальные возможности, предоставляемые LLM, для решения задач, обучения, а также для учета прошлых взаимодействий и работы с инструментами.

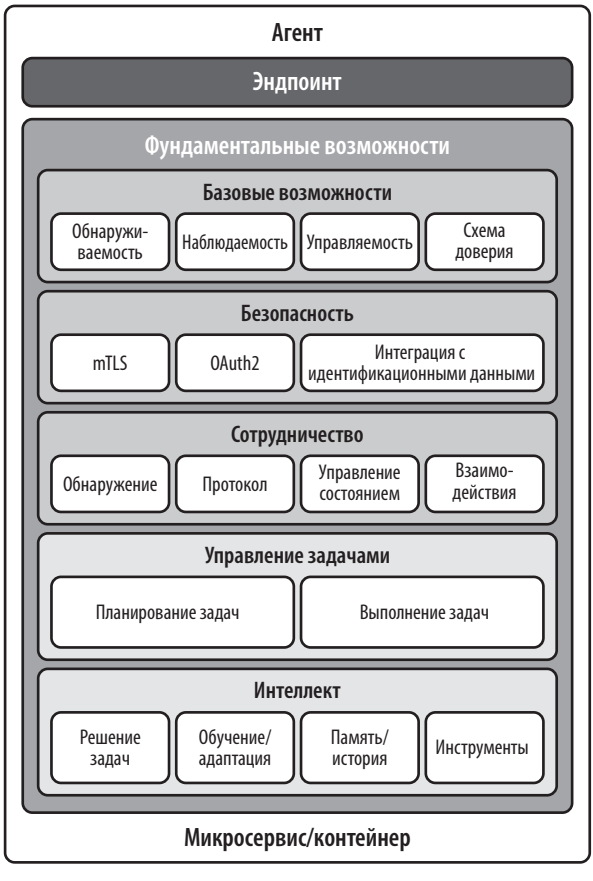


Рис. 1.2. Агенты корпоративного уровня

Агентная сетка: экосистема агентов

Мы полагаем, что эти потребности корпоративного уровня будут удовлетворены и скоро агенты начнут стремительно развиваться. Однако при этом они не будут работать в изоляции, а будут образовывать взаимосвязанные экосистемы как внутри предприятий, так и, возможно, между отраслями и предметными областями.

Эта взаимосвязанная экосистема, изображенная на рис. 1.3, и является тем, что мы называем агентной сеткой. Цель агентной сетки и ее проектирования проста:

создать условия, в которых агенты могут легко находить друг друга, безопасно сотрудничать, взаимодействовать и совершать транзакции.

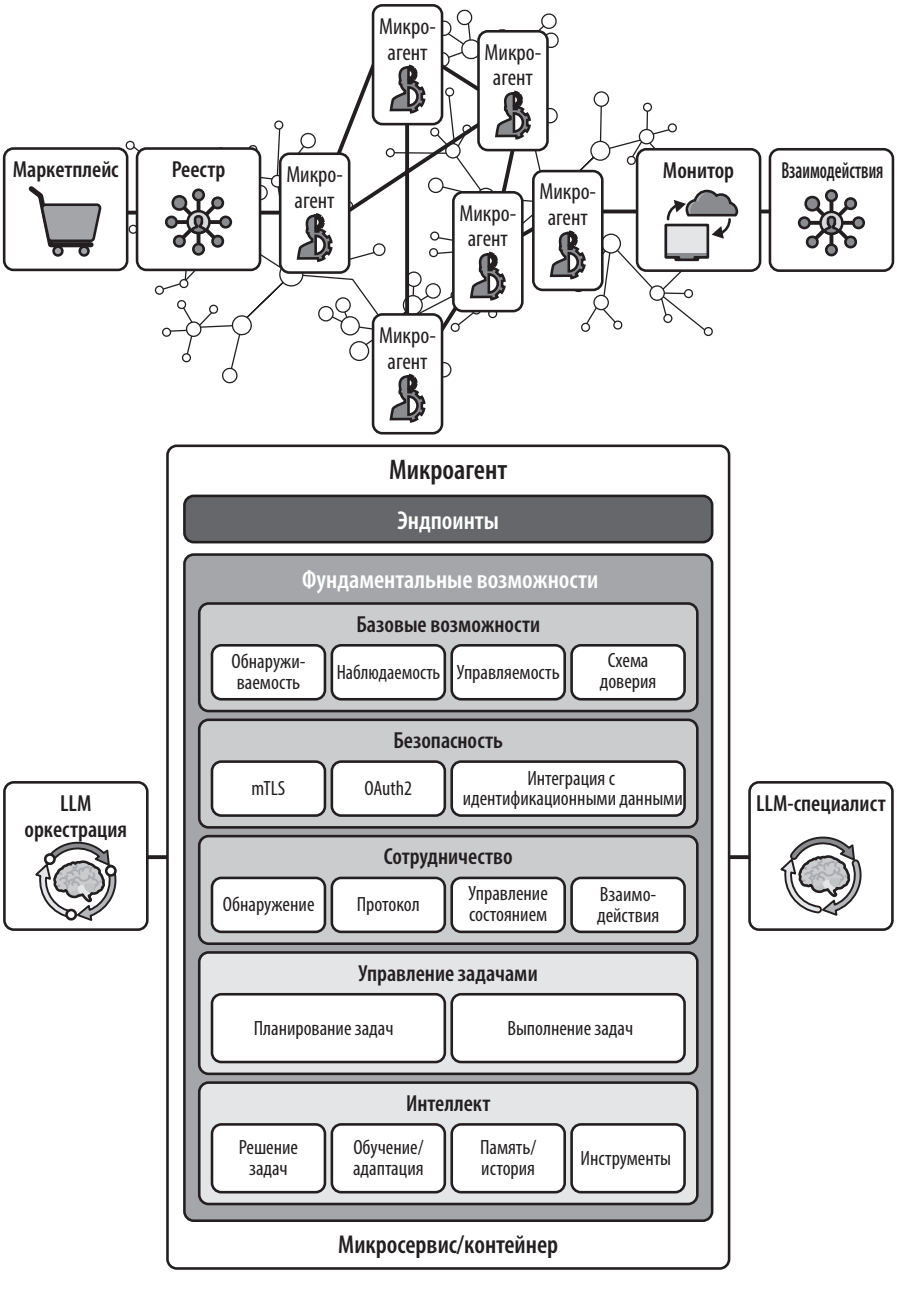


Рис. 1.3. Агентная сетка как экосистема агентов

Агенты — по крайней мере их будущие воплощения — будут действовать практически так же, как люди, несмотря на то что агенты реализованы на программном уровне и взаимодействуют с использованием сетей и API.

Как и люди, агенты не существуют в одиночку, а процветают в сообществе. У людей есть политики и правила, управляющие их поведением; есть они и у агентов. Как и у людей, никакой агент в одиночку не справится с серьезным испытанием, и, как и люди, агенты должны работать в командах для решения больших и сложных задач. Люди объединяются в многоуровневые группы — *«правительства»* — для создания законов, политик и нормативов; скоро агенты начнут создавать похожие управляющие структуры. Люди организуются в экосистемы, которые мы называем *«бизнесами»*, для предоставления услуг, которые отдельный человек не сможет предоставить в одиночку; скоро то же самое будут делать агенты.

Агентная сетка — экосистема агентов — предоставляет услуги, которые позволяют агентам сотрудничать. Она обеспечивает схему доверия (trust framework), чтобы ими можно было управлять, а также систему безопасности для их защиты. Она предоставляет основу (например, микросервисы), которая позволяет агентам работать эффективно и на корпоративном уровне.

Как нетрудно догадаться по названию, агентная сетка (agentic mesh) — агентная экосистема корпоративного уровня — является основной темой этой книги. Она состоит из шести компонентов.

Маркетплейс

Маркетплейс позволяет людям узнавать об агентах и взаимодействовать с ними. Через маркетплейс пользователи могут найти агентов, соответствующих их конкретным потребностям, а после этого инициировать взаимодействия с нужным из них.

Менеджер взаимодействий

Управляет взаимодействиями с агентами и историей и предоставляет средства, при помощи которых люди могут обратиться к агентам для выполнения задач. Также позволяет наблюдать за действиями агентов, понимать статус задач и убеждаться в том, что они выполняются в соответствии с ожиданиями и политиками.

Реестр

Реестр агентов действует как репозиторий для метаданных агентов. Они включают такую важную информацию, как назначение агента, информация о владельце, политики, роли безопасности, функциональность, описания эндпоинтов и состояния жизненного цикла. Поддерживая эти метаданные, реестр позволяет агентам через простые механизмы запросов находить друг друга, определять возможных сотрудников и, как следствие, предоставлять структурированную и безопасную среду для агентных операций.

Монитор

Монитор агентов поддерживает операционные метрики (количество запросов, задержка и т. д.) и предоставляет доступ к подробностям взаимодействий «агент/агент» и «человек/агент», истории планов, использованных агентами для выполнения задач, и другой статистике. Монитор агентов также предоставляет интерфейс для безопасного получения этой информации участниками экосистемы.

Схема доверия

Агентная сетка также имеет схему доверия, которая сертифицирует поведение агентов. Как упоминалось ранее, у каждого агента имеется описание и назначение, которые определяют защитные ограничения для его поведения. Каждый агент также поддерживает набор политик, которые закрепляют его ожидаемые цели и мотивации, а также функцию аттестации, которая определяет соблюдение агентом его политик. Схема доверия предоставляет формальную сертификацию вроде тех, которые стандартизирующие организации предоставляют сегодня (например, Underwriters Laboratories в США или Канадская ассоциация стандартов для промышленных товаров) для подтверждения того, что поведение агента соответствует ожиданиям.

Паттерны и протоколы

Агентная сетка также предлагает набор паттернов и протоколов, позволяющих агентам находить друг друга и безопасно взаимодействовать, сотрудничать и выполнять операции. Она определяет методы доступа, параметры и структуру сообщений, с которыми каждый агент будет обладать обнаруживаемостью, наблюдаемостью и управляемостью в экосистеме агентов.

Проблемы при использовании агентов

Как и в случае с любой другой революционной технологией, было бы неправильно не упомянуть, что распространение автономных агентов не обходится без проблем. Они сопоставимы с возможностями, которые открывают агенты, а может быть, даже превышают их, отражаясь на критически важных аспектах нашей экономики, общества и даже наших этических представлений.

В блоге Международного валютного фонда (МВФ) Кристалина Георгиева (Kristalina Georgieva) предупреждает, что «ИИ повлияет почти на 40 процентов рабочих мест по всему миру, заменяя одни и дополняя другие... Мы стоим на грани технологической революции, которая может обеспечить резкий прирост производительности, способствовать глобальному росту и поднять доходы по всему миру. Тем не менее она также может привести к замене рабочих мест и углублению неравенства».

Этот жесткий прогноз подчеркивает обоснованные опасения: хотя агенты обладают потенциалом для обеспечения роста и производительности, они также

создают риски для занятости и экономического равенства. Последствия не будут однородными. Отрасли, регионы и демографические группы могут столкнуться с совершенно разными последствиями: одни окажутся в выигрыше, а другие понесут потери.

Но более масштабные последствия выходят за рамки рабочих мест и производительности; агенты ставят под сомнение традиционные концепции подотчетности и надзора. Когда агенты действуют независимо, принимая решения и обучаясь на ходу, разумно спросить, кто в итоге несет ответственность за их действия? Этот вопрос распространяется на юридические, этические и операционные аспекты. Как убедиться в том, что агенты будут действовать в намеченных границах? Что произойдет, если они не справятся с задачей или принесут вред? Баланс инноваций с защитными мерами требует беспрецедентного сотрудничества между разработчиками политик, отраслевыми лидерами и технологами.

Этические факторы не менее важны. Когда агенты принимают роли, включающие принятие решений, в их действиях могут отражаться смещения и ограничения данных и алгоритмов, на которых они базируются — фактически они могут распространять свои смещения и недопонимания. Без пристального внимания это может привести к непредвиденным последствиям, включая укрепление вредных стереотипов или принятие решений, расходящихся с человеческими ценностями. Очень важно с самого начала встроить в агентов прозрачность, справедливость и подотчетность.

Эти проблемы дополнительно усугубляются скоростью, с которой развиваются агентные технологии. Агенты базируются на LLM, способности которых экспоненциально растут, а затраты на них столь же стремительно снижаются. Нравится нам это или нет, агенты приходят в наш мир. К сожалению, правительства, организации и компании с большой вероятностью будут играть в догонялки, стараясь соорудить политики и стратегии, которые бы не отставали от инноваций.

Однако при всех проблемах очевидно: нужно что-то делать. Потенциал агентов слишком велик, чтобы его игнорировать, а их интеграция в общество уже началась. Если подойти к новой эпохе осмотрительно и с готовностью адаптироваться, возможно, нам удастся освоить потенциал агентов при сокращении связанных с ними рисков. Как бы то ни было, придется найти правильное решение, иначе проблемы могут выйти из-под контроля.

Перспективы использования агентов

Современные компании уже сейчас изучают возможности внедрения ИИ в свои процессы и даже во взаимодействия с клиентами. А что касается агентов, несмотря на их относительную новизну, компании ищут способы, где и как их можно интегрировать. Да, эти стратегические обсуждения ведутся на высоком уровне, но организации по всему миру сейчас вкладываются в проверку жизнеспособности концепции и реализацию агентов, и сейчас они находятся на ранней стадии развертывания для эксплуатации. Понимание агентов и агентной экосистемы,

в которой они работают, вскоре станет одной из основных компетенций современного бизнес-управленца.

Для архитекторов агенты становятся новым компонентом в корпоративной сфере. Современные агенты и инструментарии, поддерживающие их работу, создают агентов с частично монолитным дизайном и использованием общей памяти, при этом почти все компоненты существуют в небольшом количестве исходных файлов Python. Проще говоря, этот монолитный подход не лучшая основа для простого создания агентов, пригодных для реальной эксплуатации. (То есть агентов корпоративного уровня; этот вопрос более подробно изложен в последующих главах, а пока будем считать, что агенты соответствуют примерно тем же ожиданиям, что и другие корпоративные приложения.) Дело архитектора — сформировать технологический ландшафт агентов, выявить базовые компоненты и интеграции, и определить характеристики, необходимые для перехода на корпоративный уровень. В этой книге описывается архитектура агентов, их связи и интеграции, и то, как агенты могут перейти на корпоративный уровень.

Для разработчиков сейчас появляются действительно инновационные агентные инструментарии. Но в наши дни создание агентов требует глубокого понимания LLM, промпт-инжиниринга и опыта программирования (иногда на нескольких языках). Отладка агентов, даже относительно простых, достаточно сложна (и это понятно). В этой книге мы определим ключевые компоненты, которые необходимо создать, чтобы упростить создание агентов.

Как говорилось ранее, агенты должны быть продуктами *корпоративного уровня*. Они должны вести себя как любая другая система в эксплуатации: быть обнаруживаемыми, наблюдаемыми и управляемыми, чтобы с ними можно было работать, как с любой другой системой. И еще агенты должны иметь операционные процессы, которые также упрощают выполнение действий с ними, отладку, диагностику и разрешение проблем. Мы предлагаем варианты (например, построение и выполнение агентов как микросервисов), которые позволяют разработчикам, инженерам и операционному персоналу применять свой многолетний опыт для построения приложений, пригодных для реальной эксплуатации.

Инженерам мы опишем эти операционные проблемы. Мы объясним, как агенты вписываются в корпоративные процессы DevSecOps, а также более новые процессы MLOps. Вы узнаете, какие защитные меры нужно будет предусмотреть, чтобы агенты были безопасными, чтобы используемые ими данные и их взаимодействия с клиентами соответствовали нормативным, юридическим и корпоративным требованиям, включая стандарты защиты персональных данных. Но даже если отойти от вопросов безопасности: если потенциал агентов реализуется хотя бы на долю того, во что верят оптимисты, необходимо, чтобы агентам можно было доверять. Мы объясняем принципы, лежащие за схемой доверия, которые делают агентов не только безопасными, но и заслуживающими доверия: они соблюдают этические нормы, прозрачны и делают то, что должны.

Во всех современных ИИ-решениях LLM используется в качестве мозга, но эти приложения не обладают встроенной функциональностью, обеспечивающей их безопасность. LLM может игнорировать политики организации, особенно если она явно не обучалась на них. Но даже LLM, которые прошли тонкую настройку на корпоративных знаниях, могут не понимать, как предотвратить непреднамеренную утечку чувствительной информации. А LLM, использующие векторные базы данных для *генерации, дополненной поиском* (Retrieval-Augmented Generation, RAG), создают эмбеддинги с использованием информации из защищенных баз данных, при этом игнорируя политики безопасности и правила, определявшие права доступа и конфиденциальность. Из-за этого профессионалы в области безопасности признают, что в настоящее время агентов защитить нелегко, и предлагают другие варианты повышения их безопасности.

Итоги

Мы надеемся, что предложенные методы создания архитектуры, проектирования и реализации для агентов и экосистемы агентной сетки помогут организациям устранить недостатки текущих подходов и инструментариев. Хочется верить, что эта книга позволит вам — практику, разработчику, менеджеру или руководителю — создавать агентов корпоративного уровня, готовых к эксплуатации (и, конечно, приносящих пользу), как и любые другие важные корпоративные системы. Надеемся, в итоге вы согласитесь, что агентные сетки выполняют свою основную цель: сформировать агентную экосистему, которая позволяет агентам легко находить друг друга и безопасно сотрудничать, взаимодействовать и совершать транзакции.

А еще мы надеемся, что, следуя этим практическим шагам, организации смогут преобразовать свой подход к использованию агентов, ускорить их принятие и осознать колоссальные возможности, которые они открывают.