Практический ИИ

ЗАДАЧА ДЛЯ ИИ-АГЕНТА

Практический справочник постановки, приёмки и экономики

Гордей Вихрев

октябрь 2026

Гордей Вихрев

Задача для ИИ-агента

«Автор»
2026

Вихрев Г.

Задача для ИИ-агента / Г. Вихрев — «Автор», 2026

Вы платите за доступ к ИИ-агенту и получаете гладкий текст, который нельзя принять в работу? Проблема не в модели, а в том, как поставлена задача и как устроен контроль. Эта книга — практический справочник для предпринимателя и специалиста, который хочет передать рутину агенту и получать результат, а не имитацию. Внутри: из чего состоит задание, которое исполняется с первого раза; четыре слоя приёмки и чек-лист на двенадцать пунктов; экономика — как посчитать стоимость принятого результата и не обмануться; границы безопасности, прав и ответственности; каталог двенадцати типовых провалов с лечением; регламент встройки в малой фирме и дорожная карта первых тридцати дней. Отдельная глава — разбор реального проекта с проверенными условиями площадок, формулами расчёта и честным списком неудач. Все числа — с датой проверки и ссылкой на первоисточник. Никакой мотивации: формы, метрики и решения, которые можно применить завтра утром.

© Вихрев Г., 2026
© Автор, 2026

Содержание

Об этой книге 6
Введение. Для кого эта книга и как её читать 7
Одна история, с которой начинается почти всё 7
Для кого 8
О методе, которым написана эта книга 9
Как читать 10
Четыре вещи, которые стоит запомнить до начала 11
Глава 1. Агент — не чат-бот: где проходит граница 12
Почему половина задач, поставленных агенту, обречена 12
Три слоя, которые важно не путать 13
Где агент силён, а где стабильно проваливается 14
Агент устойчиво силён там, где 14
Агент устойчиво проваливается там, где 14
Граница автономности: четыре уровня 15
Что делать с этим прямо сейчас 16
Глава 2. Язык, на котором агент понимает: анатомия задания 17
Почему обычные технические задания не работают 17
Блок 1. Результат, а не действие 18
Блок 2. Ограничения и запретительный список 19
Блок 3. Образец и антиобразец 20
Блок 4. Порядок и контрольные точки 21
Блок 5. Критерий приёмки и способ доказательства 22
Блок 6. Ресурсы и цена ошибки 23
Скелет задания: шаблон, который можно копировать 24
Две вещи, которые важнее шаблона 25
Глава 3. Приёмка: как понять, что результат настоящий 26
Главная ловушка: гладкость воспринимается как правильность 26
Четыре слоя приёмки 27
Слой 1. Формальная валидность (автоматически, 100 % покрытие) 27
Слой 2. Фактическая проверка (выборочно, но по правилам) 27
Слой 3. Логическая связность (чтение с карандашом) 28
Слой 4. Вкус и польза (последний, только после первых трёх) 28
Приёмочный чек-лист: универсальная заготовка 29
Конец ознакомительного фрагмента. 30

Гордей Вихрев
Задача для ИИ-агента

Задача для ИИ-агента

Практический справочник постановки, приёмки и экономики

Составитель — предприниматель, текст подготовлен ИИ-агентом под редакцией составителя

Гордей Вихрев

© Гордей Вихрев, 2026. Составление, отбор материала, структура и редакция — составитель.

Книга носит информационно-практический характер и не является юридической, налоговой или кадровой консультацией. Все приведённые ставки, комиссии, лимиты и сроки указаны на дату проверки — 26 октября 2026 года — и взяты из открытых официальных документов площадок; перед принятием решений проверяйте их актуальность у первоисточника. Расчёты в книге — модель, а не гарантия коммерческого результата: они показывают порядок величин и способ мышления, а не обещанный доход. Названия компаний и площадок упоминаются исключительно в описательных целях, без рекламных обязательств.

Об этой книге

Вы платите за доступ к ИИ-агенту и получаете гладкий текст, который нельзя принять в работу? Проблема не в модели, а в том, как поставлена задача и как устроен контроль. Эта книга — практический справочник для предпринимателя и специалиста, который хочет передать рутину агенту и получать результат, а не имитацию. Внутри: из чего состоит задание, которое исполняется с первого раза; четыре слоя приёмки и чек-лист на двенадцать пунктов; экономика — как посчитать стоимость принятого результата и не обмануться; границы безопасности, прав и ответственности; каталог двенадцати типовых провалов с лечением; регламент встройки в малой фирме и дорожная карта первых тридцати дней. Отдельная глава — разбор реального проекта с проверенными условиями площадок, формулами расчёта и честным списком неудач. Все числа — с датой проверки и ссылкой на первоисточник. Никакой мотивации: формы, метрики и решения, которые можно применить завтра утром.

Введение. Для кого эта книга и как её читать

Одна история, с которой начинается почти всё

Знакомая картина. Предприниматель слышит, что ИИ экономит часы. Платит за доступ. Ставит задачу своими словами — теми самыми, которыми поставил бы её живому помощнику. Получает текст. Читает. Кивает: написано складно, по делу, местами даже лучше, чем он сам бы сформулировал. Откладывает. Через неделю вспоминает, что применение так и не состоялось.

Если присмотреться, ничего плохого не произошло: текст хороший, время потрачено не грубо, модель не подвела. Произошло другое — результат получился, но не стал делом. И это, а не «плохой ИИ», объясняет, почему через полгода после начала разговоров об автоматизации у большинства малых фирм не изменилось ровным счётом ничего.

Эта книга не про то, как писать хорошие тексты с помощью нейросетей. Она про то, как устроить производство, на котором чужая — машинная — работа превращается в ваш результат: принятый, проверенный, вложенный в дело.

Для кого

Для владельца микрофирмы или самозанятого, у которого всё держится на личных руках, и он интуитивно чувствует: рутину пора отдавать, но непонятно, кому и как, чтобы не получить лишнюю головную боль. Вам не нужны технические познания, но нужны честность по поводу своих критериев и готовность два часа в неделю тратить на контроль.

Для руководителя проекта или отдела, который отвечает за результат перед людьми и деньгами. Вам важна не магия, а управляемость: метрики, границы ответственности, процедура на случай сбоя.

Для специалиста, который хочет перестать быть рукой и стать головой: делегировать исполнение и оставить себе суждение. Это самая благодарная аудитория, потому что у вас уже есть критерии качества — вы просто не умеете их формулировать для исполнителя, у которого нет вашего опыта.

Не для вас книга, если вы ищете способы обойти требования площадок, сгенерировать тысячу текстов на продажу или найти «пассивный доход без усилий». Шестая глава объяснит, почему это не работает, и с цифрами.

О методе, которым написана эта книга

Честность требует сказать главное: **этот текст подготовлен ИИ-агентом.** Под редакцией, с постановкой задач, проверкой фактов по первоисточникам, с отказом от тем, которые не удалось подтвердить, и с разбором собственных неудач. Я не скрываю это и не извиняюсь — наоборот, книга построена так, чтобы вы могли проверить каждый шаг.

Почему это важно упомянуть в начале. Потому что основная опасность, о которой я пишу в третьей и седьмой главах, — это как раз способность машин производить правдоподобный текст. Читатель, который не знает, кто автор, не сможет применить книгу правильно: он будет искать у автора мудрость, тогда как ему нужен метод, который он проверит сам.

Что здесь сделано человеком: отбор тем, структура, решения о том, что выбросить, проверка чисел по официальным документам площадок, отказ от целой главы (про налоги), потому что первоисточники не открылись, и финальная редакция. Это не формальность: именно так выглядит «существенная переработка», которую требуют правила приёма у всех серьёзных площадок, и именно поэтому в книге есть глава 6 с подлинными документами, а не глава 6 с общими рассуждениями.

Как читать

Книга написана так, чтобы её можно было использовать, а не только прочитать. Три режима.

Режим первый — сквозной. Главы 1–5 читаются подряд: они строят каркас понимания (границы, задание, приёмка, экономика, безопасность). Пропускать нельзя: каждая следующая опирается на предыдущие, и попытка начать с середины даёт типичное ощущение «это я и так знал» без изменения практики.

Режим второй — справочник. Глава 7 (провалы), глава 8 (выбор исполнителя) и приложение — это стол, к которому возвращаются. В приложении — пустые формы, заполненный пример, чек-листы, таблица «симптом → лечение», словарь. Если вы уже что-то пробовали и у вас что-то не получилось, начните с таблицы в приложении: по симптому вы найдёте провал и лечение за две минуты.

Режим третий — план. Глава 10 — это не теория, а расписание на месяц. Если вы хотите результат, а не впечатление, откройте её прямо сейчас, сделайте день 1–2 (список задач с тремя метками) и вернитесь к остальным главам по мере надобности. Так читают деловые книги те, кто что-то меняет, а не те, кто получает удовольствие от чтения.

Четыре вещи, которые стоит запомнить до начала

Агент исполняет формулировку, а не намерение. Всё, что вы не описали, будет описано за вас.

Считать надо стоимость принятого результата, а не стоимость запроса. Иначе экономика выглядит сказкой, а оказывается расходом.

Автоматизируйте производство, но не автоматизируйте доверие. Приёмка — это и есть та часть, которую нельзя отдать.

Прочтите правила до начала работы. Половина провалов в главе 6 случилась бы иначе, если бы этот совет применили заранее.

И последнее. Эта книга не обещает, что автоматизация сделает вас богаче. Она обещает ровно одно: если вы примените описанные формы и метрики, вы будете знать о своём процессе больше, чем знаете сейчас, — в числах, а не в ощущениях. А это, как показывает практика, то, с чего начинаются любые реальные изменения.

Глава 1. Агент — не чат-бот: где проходит граница

Почему половина задач, поставленных агенту, обречена

Начну с наблюдения, которое дороже любой методики. Большинство владельцев малого бизнеса, впервые открывая ИИ-агента, ставят ему задачу в той же форме, в какой поставили бы её живому сотруднику. «Проанализируй рынок и предложи, на чём заработать». «Сделай так, чтобы пошёл доход». «Разберись с документами». И затем удивляются результату: длинному, гладкому, уверенному тексту, который нельзя ни принять, ни использовать.

Проблема не в модели. Проблема в том, что живому сотруднику такую задачу поставить можно — потому что он придёт и задаст уточняющие вопросы. Он спросит: а какой бюджет? а что уже пробовали? а что для вас значит «доход»? Агент же уточняющих вопросов, как правило, не задаёт: он добирает недостающее правдоподобной гипотезой. И чем сильнее модель, тем убедительнее выглядит догадка. Отсюда главный парадокс автоматизации: **лучший исполнитель оказывается худшим собеседником**.

Запомните одну фразу, которая экономит десятки часов: агент исполняет не намерение, а формулировку. Ваше намерение он угадывает, а формулировку — выполняет буквально и с удовольствием.

Три слоя, которые важно не путать

Разговор про «ИИ в бизнесе» тонет, потому что под одним словом прячут три разные вещи.

Первый слой — генератор. Модель, которая по запросу выдаёт текст, картинку, код. Она ничего не помнит между запросами, ничего не проверяет и ни за что не отвечает. Генератор полезен, когда нужен черновик, вариант, идея. Он бесполезен, когда нужен результат.

Второй слой — агент. Это генератор, которому дали инструменты: файловую систему, сеть, терминал, возможность запускать программы и читать ответы этих программ. принципиальное отличие не в «интеллекте», а в **наличии обратной связи**. Агент может написать скрипт, запустить его, увидеть ошибку, исправить и запустить снова. Именно это делает возможной автоматизацию там, где раньше требовался глаз и рука.

Третий слой — процесс. Набор агентов, расписаний, правил приёмки и людей, которые ставят цели. Процесс — это то, чем на самом деле управляет предприниматель. Ошибка новичка — купить второй слой и ожидать, что появится третий. Третий слой не появляется сам: он проектируется.

Полезная проверка перед тем, как что-то автоматизировать: ответьте, на каком слое вы формулируете задачу. Если на втором («агент, сделай сайт») — вы получите второй слой. Если на третьем («вот цель, вот ограничения, вот критерий приёмки, повторяем еженедельно») — появляется шанс на результат.

Где агент силён, а где стабильно проваливается

Ниже — карта, собранная не из теории, а из наблюдений за реальными прогонами. Разделение грубое сознательно: тонкие градации в этом вопросе вредят, потому что позволяют принять «иногда работает» за «работает».

Агент устойчиво силён там, где

Есть машинная проверка. Задача считается выполненной, если проходит тест, валидацию, компиляцию, сходится по контрольной сумме. Код, разметка данных, форматы файлов, математические расчёты, преобразование одного структурированного представления в другое. Здесь агент работает почти без человеческого участия: он сам видит, прав ли он.

Есть исчерпывающий источник рядом. Если ответ лежит в документе, который агент может скачать и прочитать, он найдёт его быстрее и внимательнее любого сотрудника. Поиск по первоисточникам, выписывание формулировок, сопоставление версий документа — рутина, которую давно пора отдать.

Нужен объём, а не изящество. Обработать четыреста карточек, перевести сто строк, разложить по полочкам архив файлов. Агент не устаёт, не раздражается и не начинает халтурить на сотой позиции. Единственное, что он принесёт в подарок, — монотонную последовательность, но об этом в главе про приёмку.

Требуется несколько вариантов. Когда нужно двадцать формулировок, чтобы выбрать одну, агент дешевле секретаря в стократ.

Агент устойчиво проваливается там, где

Критерий успеха живёт у вас в голове. «Сделай красиво», «чтобы цепляло», «нормальный вариант». Агент подставит своё среднее понимание нормы — и получит ровно его, уверенно и с аргументацией. Единственное лечение: перевести критерий в измеримый вид до постановки задачи. Не можете измерить — не автоматизируйте, сначала делайте руками и фиксируйте, по каким признакам принимали.

Ошибка дорога и необратима. Удаление данных, отправка писем контрагентам, публикация на публичных ресурсах, финансовые операции, любые действия «вживую». Здесь нужен не агент, а агент с подтверждением: он готовит, человек нажимает. Экономия на этом правиле всегда оборачивается дороже.

Нужна ответственность перед третьими лицами. Налоговая отчётность, юридические выводы, медицинские рекомендации, обещания клиенту по срокам. Агент не несёт ответственности, и никакой формулировкой в промпте это не изменить. Решение принимается человеком, который ставит подпись.

Факты не проверяемы из доступного источника. Это самая коварная зона. Модель знает «в среднем правильно» и формулирует это с той же интонацией, что и проверенный факт. Я специально начал эту книгу с отказа от налоговой тематики: пока не смог открыть первоисточники напрямую, я не имел права писать главу про АУСН. Хороший агент обязан сказать «не нашёл подтверждения» — но он скажет это только если вы разрешили ему признавать незнание. Об этом — в главе про приёмку.

Граница автономности: четыре уровня

Практичный инструмент, которого нет в популярных инструкциях. Прежде чем давать задачу, решите, на каком уровне работаете. Уровни нумеруются по возрастанию риска.

Уровень 0 — консультация. Агент только читает и объясняет. Ничего не меняет в вашем мире. Цена ошибки: зря потраченное время.

Уровень 1 — черновик. Агент создаёт артефакт в изолированной папке, который не попадает к людям без вашей проверки. Цена ошибки: переделка. Это рабочий уровень для 80 % задач малого бизнеса, и именно с него надо начинать всё, что не научились принимать.

Уровень 2 — исполнителя с самопроверкой. Агент выполняет действие и сам доказывает, что оно правильное: запустил тест, сравнил контрольную сумму, открыл результат и перечитал. Цена ошибки: обнаруживается не сразу, но обнаруживается. Требует, чтобы вы заранее описали проверку, а не только желаемый результат.

Уровень 3 — автономный с внешним контуром. Агент действует сам, но вокруг него стоят ограничители: лимит сумм, белый список операций, журнал всех действий, обязательный отчёт. Цена ошибки: контролируемая, если ограничители спроектированы заранее, и катастрофическая, если «потом допилим».

Уровень 4 — без контура. Полная свобода. Его не существует как разумного выбора. Если кто-то предлагает уровень 4 — он продаёт вам страх, а не технологию.

Разница между новичком и опытным пользователем агентов почти всегда в одном: новичок пытается прыгнуть на уровень 3, потому что «ради чего тогда всё это». Опытный держит уровень 1–2 и наращивает долю автоматизированных задач, а не глубину автономии. Доля задач растёт быстро, если не терять контроль.

Что делать с этим прямо сейчас

Прежде чем читать дальше, сделайте одну вещь, которая окупается немедленно: выпишите десять задач, которые вы делали руками за последнюю неделю и которые были вам неприятны. Против каждой поставьте метку:

Есть ли машинная проверка результата? (да/нет)

Лежит ли ответ в доступном источнике? (да/нет)

Ошибка обратима? (да/нет)

Задачи, где на все три вопроса ответ «да», — ваш первый контур автоматизации. Их можно отдавать агенту на уровень 2 почти без риска. Задачи, где «нет» на третий вопрос, — только через подтверждение человеком, сколько бы времени это ни отнимало. Задачи, где «нет» на первый, — сначала научитесь принимать руками и формулировать критерий, потом автоматизируйте.

Этот простой фильтр даёт больше, чем большинство методик внедрения: он запрещает начинать не с того. А начинать не с того — главная причина, по которой малый бизнес через полгода говорит: «Попробовали мы вашего ИИ, толку нет».

Глава 2. Язык, на котором агент понимает: анатомия задания

Почему обычные технические задания не работают

Техническое задание для человека и задание для агента — это разные жанры литературы, хотя все привыкли думать, что это один жанр.

Техническое задание человеку пишется на языке намерений и соглашений: «сделай страницу товара в нашем фирменном стиле», «подготовь отчёт по продажам», «надо, чтобы вызывало доверие». Человек дополняет пробелы здравым смыслом, опытом и — что важнее всего — вопросом «а что ты имеешь в виду под фирменным стилем?». Пробелы в задании человеку не вредят: они заполняются общей картирой мира.

Задание агенту на том же языке приводит к обратному результату. Пробел заполняется статистической средней по тому, что модель видела чаще всего. Среднее — это не ваш вкус, не ваш бизнес и не ваша ситуация. Оно просто правдоподобно.

Отсюда практический вывод, который стоит сформулировать максимально плоско: **в задании агенту нельзя оставлять пробел, если вы не готовы принять любое его заполнение.** Всё, что вы не описали, будет описано за вас — и, скорее всего, не так, как вы держали в голове.

Хорошее задание агенту состоит из шести блоков. Я называю их «каркас», потому что они не зависят от темы задачи: каркас один, наполнение разное.

Блок 1. Результат, а не действие

Первая и самая частая ошибка — описывать процесс вместо результата. «Проанализируй конкурентов» — это процесс. Он не имеет конца, критерия и приёмки. Агент будет анализировать ровно столько, сколько позволит ему контекст, и остановится с видом человека, выполнившего работу.

Правильная форма: «Полученный файл содержит список из не менее чем двенадцати конкурентов, по каждому — цена, позиция, три отличия от нашего предложения; файл лежит по пути …; формат — таблица с этими колонками …».

Проверка на пригодность: можно ли написать программу, которая определяет, выполнено ли задание? Если да — задание сформулировано как результат. Если нет — вы описали процесс, и приёмки не будет.

Блок 2. Ограничения и запретительный список

Ограничения важнее указаний. Указание сужает пространство решений до приятного, запрет — отрезает опасное. Модель охотнее нарушает то, о чём ей не сказали, чем то, что ей запретили явно.

Запретительный список пишется в отрицательной форме и конкретно:

не выдумывать факты; при отсутствии подтверждённого источника — писать «не проверено» и указывать, что именно требовалось проверить;

не трогать файлы вне указанного каталога;

не отправлять ничего вовне (письма, публикации, платежи) без отдельного подтверждения;

не использовать общие формулировки вроде «многие эксперты считают», «как правило», «в современном мире»;

не сокращать объём, если он не согласован; не раздувать его ради вида полноты.

Пятый пункт заслуживает пояснения. Агенты имеют привычку к «воде» — не из злого умысла, а потому что гладкая связка дешевле в генерации, чем плотный факт. Запрет на воду даёт прибавку качества больше, чем любое улучшение формулировки основного требования.

Блок 3. Образец и антиобразец

Самый мощный блок и самый редко используемый. Модель понимает образец лучше, чем любое описание. Одно пример-эталон объёмом в двести слов заменяет страницу требований и работает точнее.

Работает схема «три образца»:

Как надо — фрагмент, которому соответствует ожидаемый результат.

Как не надо — типичная ошибка, которую вы уже видели. Подпись к нему обязательна: «здесь плохо вот это, потому что…».

Граница — случай, где вы сами не уверены, и как он должен быть обработан: «если не можешь отличить от антиобразца — остановись и спроси».

Третий образец важнее первых двух: он превращает неопределённость из источника ошибок в источник вопросов. Без него агент будет уверенно делать неверное. С ним — хотя бы признается.

Формулировка границы имеет специфический вид: «Если требование противоречит другому требованию или данные не сходятся, — не выбирай сам, а зафиксируй противоречие и предложи два варианта решения с ценой каждого».

Блок 4. Порядок и контрольные точки

Для задач длиннее получаса работы нужна разбивка с промежуточной приёмкой. Логика простая: чем позже обнаружена ошибка, тем дороже её исправление. В задачах с агентом ошибка, обнаруженная в конце, обычно означает не исправление, а переделку всего — потому что агент не умеет «переделать половину», он переделывает.

Контрольные точки формулируются как события остановки:

после плана: показать структуру и получить подтверждение;

после первой единицы результата (одна страница, один раздел, одна карточка): показать и получить подтверждение, что направление верное;

после 30 % объёма: показать метрики и риски отставания;

после 100 % объёма: показать самопроверку по чек-листу.

Точка «после первой единицы» — самая выгодная инвестиция в этой главе. Она стоит один лишний цикл общения и спасает от ситуации, когда через три часа вы получаете 150 страниц не в ту сторону.

Блок 5. Критерий приёмки и способ доказательства

Критерий приёмки — это не «мне должно понравиться». Это список проверяемых утверждений о результате, каждое из которых может быть истинно или ложно.

Пример плохого критерия: «Текст должен быть грамотным и полезным». Пример хорошего:

объём — от 120 000 знаков с пробелами;

ни одно число в тексте не появилось без источника с датой проверки;

в тексте нет утверждений о законах, если первоисточник не открыт и не процитирован;

средний размер абзаца — от 4 до 9 строк, врезок без текста не более трёх подряд;

каждый раздел заканчивается практическим выводом объёмом не менее трёх предложений;

автоматическая проверка формата пройдена без ошибок.

Второе требование к блоку: **доказательство**. Агент обязан приложить к результату не только файл, но и отчёт проверки: что запускал, что сравнил, что не прошло, что осталось спорным. Без отчёта приёмка превращается в ваше личное чтение, а личное чтение — это там, где начинается человеческая ошибка: глаза скользят по гладкому тексту и принимают правдоподобие за истину.

Блок 6. Ресурсы и цена ошибки

Два последних элемента, которые обычно забывают.

Ресурсы: что агенту разрешено использовать. Конкретные пути к файлам, списки доменов, доступные инструменты, лимиты времени и денег. «Используй только эти пять источников, при недоступности одного — зафиксируй и продолжай по остальным» избавляет от двух бед: от выдумывания и от остановки на полпути.

Цена ошибки: явное указание, что дороже — потерять время или получить неточность. Это не мелочь, это переключатель поведения. Если вы не написали, агент выберет скорость: он оптимизирует то, что вы не защитили. Формулировка: «При конфликте между полнотой и скоростью приоритет — полнота; неполноту помечай, скорость не экономь на проверках».

Скелет задания: шаблон, который можно копировать

Ниже — каркас, который я использую как отправную точку для всех задач. Он не красив, он работает.

ЦЕЛЬ: <что должно появиться в мире и для чего>
РЕЗУЛЬТАТ: <файл/действие, путь, формат, минимальный объём>
ПРОВЕРКА: <как автоматически доказывается, что результат верен>
ОГРАНИЧЕНИЯ:
- <что нельзя делать ни при каких обстоятельствах>
- <что нельзя менять>
РЕСУРСЫ: <файлы, адреса, инструменты, лимиты>
ОБРАЗЕЦ: <фрагмент «как надо»>
АНТИАБРАЗЕЦ: <фрагмент «как не надо» + почему>
ГРАНИЦА: <когда остановиться и спросить>
КОНТРОЛЬНЫЕ ТОЧКИ: <где остановиться для подтверждения>
ПРИОРИТЕТ: <что важнее: полнота / скорость / точность>
ОТВЕТСТВЕННОСТЬ: <что требует подтверждения человеком>
ОТЧЁТ: <что приложить к результату, кроме самого результата>

Десять строк. На заполнение уходит три-пять минут. Экономия — от тридцати минут до нескольких часов на переделках, и это не считая случаев, которые не всплывают в удачных историях внедрения: когда гладкий неверный результат приняли и понесли дальше.

Две вещи, которые важнее шаблона

Первая: задание — это документ, а не сообщение. Если задание оказалось рабочим, сохраните его рядом с результатом. Через месяц вы захотите сделать похожую вещь и не вспомните, что именно сработало. Агенты не помнят между сессиями; ваша база заданий — это и есть накапливаемая интеллектуальная собственность, часто более ценная, чем сами результаты.

Вторая: формулировка проверяется на дешёвом случае. Прежде чем отдавать агенту большую задачу по новому шаблону, дайте ему по этому шаблону крошечную задачу: одну страницу, один расчёт, одну карточку. Если на малом объёме результат не устраивает — на большом он точно не устроит, просто вы заплатите больше. Проверка шаблона на дешёвом случае стоит две минуты и спасает дни.

Глава 3. Приёмка: как понять, что результат настоящий

Главная ловушка: гладкость воспринимается как правильность

Есть свойство человеческой психики, на котором держится большинство провалов с ИИ, и знать его обязательно, прежде чем принимать первый результат. Мы оцениваем текст по беглости чтения, а беглость по неписаной внутренней лицензии считается признаком истинности. Текст, который легко читается, кажется заслуживающим доверия. Текст шероховатый — подозрительным.

Модели сняли с этой связи всякий смысл: они пишут максимально бегло независимо от того, истинно содержание или нет. Более того, чем увереннее и гладче сформулирован неверный тезис, тем выше шанс, что он пройдёт приёмку. Отсюда практическое правило, которое стоит записать и не нарушать:

Приёмка результата агента начинается с поиска того, что он не сделал, а не с чтения того, что он сделал.

Чтение — второй этап. Первый этап — проверка границ.

Четыре слоя приёмки

Слои идут от дешёвых к дорогим. Первые три автоматизируются и снимают 90 % риска. Четвёртый — человеческий, и он должен доставаться уже чистому результату.

Слой 1. Формальная валидность (автоматически, 100 % покрытие)

Самый недооценённый и самый дешёвый слой. Результат должен быть технически пригоден: файл открывается, формат соответствует стандарту, структура цела, кодировка верная, обязательные поля заполнены.

Я приведу конкретный пример из собственной практики — ровно тот, который случается у каждого, кто автоматизирует производство документов. Перед тем как писать эту книгу, я собрал производственный конвейер: один исходник превращается в четыре формата (текстовый документ, электронная книга, файл для печати, обложка). Первый прогон дал ошибку в самом конце цепочки — файл для печати не собирался из-за неверно указанной библиотеки работы с изображениями. Если бы я принял работу по внешнему виду — три формата из четырёх были бы готовы, папка выглядела бы заполненной, отчёт был бы бодрый: «собрано». По факту четвёртый, самый важный для читателя формат отсутствовал.

Формальная валидность ловит именно это. Наборы проверок, которые стоит иметь в каждом проекте:

файл существует, ненулевой, открывается програмно;

структура соответствует ожидаемой (для архивных форматов — обязательные элементы на месте, для таблиц — колонки, для текста — заголовки);

кодировка и шрифты: кириллица не превратилась в «кракозябры», шрифты встроены, если это файл для печати;

объём в заданных границах (не меньше минимума и не раздут);

отсутствие служебных артефактов: остатков разметки, заглушек, слов-маркеров вроде «продолжение следует», «вставьте сюда», «TODO».

Эти проверки пишутся один раз на проект и работают бесплатно вечно. В моей практике они заняли около часа и поймали с десяток дефектов, которые глаз бы не увидел: несоответствие пропорций обложки требованиям площадки, отсутствующие файлы в структуре электронной книги, неверный порядок элементов архива, из-за которого часть читалок отказывается открывать файл.

Слой 2. Фактическая проверка (выборочно, но по правилам)

Здесь проверяется то, что агент утверждает. Правило простое и жёсткое: **проверяется не всё, а всё важное плюс случайная выборка.**

«Всё важное» определяется до приёмки: числа, даты, имена, цитаты, ссылки, суммы, обещания. Если в результате двадцать числовых значений, а вы проверили два — вы не провели приёмку, вы создали видимость приёмки.

Случайная выборка нужна для одной цели: обнаружить систематическую подделку. Модель, которая выдумала один факт, с высокой вероятностью выдумала и другие, а выборка — единственный способ это увидеть, не читая всего.

Отдельная категория — ссылки. Проверять надо не «ссылка открывается», а «ссылка подтверждает именно то, ради чего приведена». Я видел и сам ловил себя на том, что ссылка ведёт на живой сайт с релевантной темой, но утверждение, которое она должна подтверждать,

на той странице отсутствует. Это самый опасный вид несоответствия: он проходит обе поверхностные проверки и разоблачается только чтением.

Слой 3. Логическая связность (чтение с карандашом)

Третий слой — то, для чего нужен человек. Но читать надо не на содержание, а на стыки:

Обещано ли в начале то, что появилось в конце? Агенты любят заявлять широкую тему и выдать узкое покрытие.

Есть ли выводы в разделах, или разделы заканчиваются перечислением?

Повторяется ли одна мысль в трёх местах разными словами? Это признак того, что объём набирался, а не писался.

Противоречит ли текст сам себе на расстоянии? Модель плохо держит противоречия дальше нескольких тысяч знаков: в главе 2 она может написать «проверяйте всё», а в главе 5 — «выборочная проверка достаточна».

Противоречия на расстоянии — визитная карточка автогенерации. Живой автор, который держит книгу в голове, ошибается реже. Поэтому перечитывайте не текст, а его утверждения: выпишите ключевые тезисы списком и просмотрите список — противоречия в списке видны мгновенно, в прозе — нет.

Слой 4. Вкус и польза (последний, только после первых трёх)

Сюда доходит уже выверенный результат. Здесь решается вопрос не «правильно ли», а «надо ли». И честно: если первые три слоя пройдены, а четвёртый говорит «не надо», — виноват не агент. Виновато задание, в котором не было ответа на вопрос «зачем».

Приёмочный чек-лист: универсальная заготовка

Чек-лист, который я применяю к любому результату агента. Первые пять пунктов убивают большинство проблем.

Результат на месте и в нужной форме. Путь, имя, формат, объём. Открыть и убедиться программно, а не по отчёту.

Самопроверка приложена. Агент показал, что именно он проверял и что получилось. Если отчёта нет — приёмка не начинается, результат возвращается.

Все утверждения с фактами имеют источник с датой. Дата обязательна: утверждение, верное в 2024-м, может быть ложным сегодня.

Нет «мёртвых» мест. Ни заглушек, ни «будет добавлено», ни пустых подпунктов, ни списков без пояснений.

Границы не переступлены. Не тронуты чужие файлы, не отправлено ничего вовне, не изменено то, что менять было нельзя.

Выборочно проверены числа и ссылки

Конец ознакомительного фрагмента.

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, купив полную легальную версию на Литрес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.