

ВЕЛИКИЙ ИИ-ОБМАН



САБИР АЛМАСОВ

Сабир Алмасов

Великий ИИ-обман

«Автор»

2026

Алмасов С.

Великий ИИ-обман / С. Алмасов — «Автор», 2026

«Освой нейросети за неделю и зарабатывай сотни тысяч». «Запусти бизнес одной кнопкой». «Наш ИИ заменит целый отдел». Почему за такими обещаниями часто скрываются дорогие курсы, бесполезные подписки и расходы, о которых молчат в рекламе? «Великий ИИ-обман» Сабира Алмасова разбирает, как нам продают надежду на лёгкие деньги. Подготовленные демонстрации, скриншоты выручки вместо прибыли, секретные промпты, сомнительные гарантии и стартапы с громкими заявлениями — каждый приём рассмотрен через конкретные вопросы и способы проверки. Вы узнаете, как оценивать возможности нейросетей, считать полную стоимость услуг, проверять продавцов и читать договоры. Отдельные главы посвящены дипфейкам, инвестиционным обещаниям и возврату денег за неудачные покупки. Книга для тех, кто хочет пользоваться ИИ с пониманием его возможностей и сохранить свои деньги.

© Алмасов С., 2026

© Автор, 2026

Содержание

Вступление. За что вы платите	6
Глава 1. Как обещание становится товаром	7
Глава 2. Что скрывается за словом «нейросеть»	9
Глава 3. Уверенный ответ и цена ошибки	11
Глава 4. Что остаётся за кадром демонстрации	13
Глава 5. Испытание, которое нельзя выиграть красивым ответом	15
Глава 6. Проценты точности и удобные знаменатели	17
Конец ознакомительного фрагмента.	18

Сабир Алмасов

Великий ИИ-обман

Великий ИИ-обман

Где нас жестко надувают инфоцыгане и стартапы

Сабир Алмасов

Вступление. За что вы платите

Вам показывают ролик: человек вводит несколько строк, получает красивую картинку, затем на экране появляется сумма заработка. Между картинкой и деньгами проходит три секунды. Из этого монтажа исчезают покупатель, переговоры, отказ, переделка и расходы. Именно в этих пропусках обычно и прячется обман.

Нейросеть действительно может помочь написать текст, подготовить варианты дизайна, разобрать документы или ускорить работу с обращениями. Но продавец курса нередко предлагает совсем другую сделку. Он продаёт уверенность, что теперь зарабатывать можно без навыков, без конкуренции и почти без труда. Стартап делает похожий ход: показывает одну удачную демонстрацию и предлагает заплатить за систему, будто её надёжность уже доказана.

Эта книга посвящена расстоянию между обещанием и проверенным результатом. Мы разберём, как из возможности сделать черновик получается обещание заменить специалиста, как выручку выдают за прибыль и как доступ к программе превращают в якобы готовый бизнес. Затем перейдём к решениям: что спросить, что проверить самостоятельно, какие расходы посчитать и когда прекратить переговоры.

Слово «инфоцыгане» в названии обозначает недобросовестных продавцов знаний и заработка. Оно не связано с этническим происхождением. В тексте важны действия продавца: какие сведения он скрывает, что обещает и чем подтверждает свои слова. Дорогой курс сам по себе не является обманом. Маленький стартап не обязан иметь собственную фундаментальную модель. Ошибка программы не доказывает мошенничества. Но ни цена, ни новизна технологии не освобождают продавца от необходимости объяснить, за что он берёт деньги.

Мы не будем пытаться угадать, какая модель окажется лучшей через год. Названия сервисов меняются быстрее, чем правила хорошей покупки. Поэтому основа книги — способы проверки, которые работают и при смене продукта. Они подходят человеку, который выбирает курс для новой профессии, специалисту, покупающему подписку, и владельцу небольшой компании, которому предлагают автоматизировать продажи.

Все придуманные ситуации и суммы обозначены как условные. Они нужны для расчёта и сравнения, а не для обещания дохода. Реальные разбирательства сопровождаются ссылками на первичные материалы в конце книги. Сведения проверены на 13 октября 2026 года. Описанные иностранные дела показывают механизм вводящих в заблуждение обещаний; российские права и процедуры рассматриваются отдельно.

Вы можете читать книгу последовательно. Если покупка уже сделана, особенно пригодятся главы о договоре и возврате. Однако не пропускайте объяснение ошибок и тестов: без него трудно сформулировать, что именно оказалось не так. Фраза «нейросеть плохая» почти ничего не сообщает продавцу, банку или специалисту, который помогает разобраться в споре. Дата, исходное задание, обещанное условие и фактический результат сообщают гораздо больше.

В конце чтения у вас должна появиться привычка задавать один короткий вопрос: какой результат я получу за эти деньги и как смогу убедиться, что получил его? Если ответ понятен, технологию можно изучать дальше. Если вместо ответа снова предлагают поверить в будущее, оплачивать это будущее необязательно.

Глава 1. Как обещание становится товаром

Рекламное обещание удобно тем, что его можно сделать гораздо шире самого продукта. Программа выдаёт варианты текста, а на странице продажи появляется «отдел маркетинга в одном окне». Бот сортирует сообщения, а в презентации он уже «полностью управляет клиентским потоком». Разрыв возникает не из-за одного преувеличенного прилагательного. Покупателю предлагают ожидать действие, которого никто ещё не проверил.

Чтобы заметить подмену, разделите сообщение продавца на три части. Первая — существующая функция: что система делает сейчас. Вторая — предполагаемый эффект: зачем это может пригодиться. Третья — ваше обязательство: сколько денег, времени и доступа к данным вы отдаёте. Реклама охотно смешивает эти части. Деньги просят сегодня, эффект относят к будущему, а в качестве доказательства показывают работающую функцию.

Условный сервис создаёт десять вариантов объявления. Это функция. Некоторые варианты могут улучшить отклик после проверки и редактуры. Это гипотеза об эффекте. Годовая подписка за 24 000 рублей — обязательство покупателя. Из наличия десяти вариантов нельзя вывести, что реклама принесёт сумму больше 24 000 рублей. Для такого вывода нужны данные о товаре, аудитории, площадке, бюджете и сравнении с прежним подходом.

Покупатель часто сам достраивает обещание. Он слышит «экономит время» и представляет свободные вечера. Слышит «создаёт видео» и представляет оплаченные заказы. Слышит «работает круглосуточно» и представляет продажи без своего участия. В договоре при этом может стоять только доступ к интерфейсу. Для продавца эта разница выгодна: широкое ожидание привлекает деньги, узкое обязательство уменьшает ответственность.

Полезно записать рекламную фразу обычным языком. «Автономный ИИ-отдел» превращается в «программа будет отвечать на такие-то обращения без предварительного одобрения человека». «Профессия за неделю» превращается в «после семи дней обучения я смогу выполнить такое-то задание для покупателя с такими-то требованиями». Если перевод не получается, обещание пока слишком расплывчато для оплаты.

Следующий шаг — найти условия. Система понимает голос, но какой язык и качество записи подходят? Создаёт дизайн, но можно ли редактировать исходник? Помогает продавать, но кому принадлежат контакты и кто отвечает на сложные вопросы? Работает без человека, но кто исправляет сбой? Условия определяют ценность гораздо точнее, чем слово «революционный».

Маркетинг может быть громким и при этом добросовестным. Честный продавец иногда использует эффектные заголовки, но способен вернуться к конкретике: показать ограничения, дать испытать функцию, назвать дополнительные расходы. Проблема начинается там, где конкретный вопрос объявляют проявлением недоверия, которое якобы мешает вам стать успешным.

Особенно внимательно относитесь к обещанию результата, зависящего от третьих лиц. Продавец подписки не управляет решениями ваших клиентов. Автор курса не управляет наймом у работодателя. Создатель торговой программы не управляет будущими ценами. Он может отвечать за качество своей услуги, но широкий внешний результат требует отдельных условий и доказательств.

Условный преподаватель предлагает научить делать короткие рекламные ролики. Он показывает программу занятий, задания и порядок обратной связи. Это обсуждаемая услуга. Если тот же преподаватель обещает каждому участнику 150 000 рублей в месяц, предмет разговора меняется. Теперь недостаточно показать красивый ролик. Нужно объяснить, откуда придут заказы, сколько человек действительно достигли результата и как посчитаны деньги.

Вы можете составить для любой покупки короткую запись: «Плачу такую-то сумму, получаю такое-то действие, проверяю его таким-то способом». Если в записи оказывается «получаю возможность попробовать», оцените цену именно возможности. Иногда она оправдана. Но её нельзя незаметно оценивать как гарантированный доход.

Отдельная ловушка — обещание будущих функций. Продукт продают по цене зрелого инструмента, объясняя недостатки тем, что скоро всё обновится. Планы развития могут быть разумными, но это не уже поставленная услуга. Спросите, что остаётся полезного, если обновление задержится на полгода. Если нынешняя версия не решает ни одной вашей задачи, вы платите главным образом за ожидание будущей разработки.

Важно также выяснить, кому продукт не подходит. Когда продавец уверен в практической ценности, он способен обсуждать случаи, в которых продукт не нужен. Например, при слишком малом числе обращений автоматизация не окупит настройку. Если инструмент объявляют одинаково необходимым всем, от студента до промышленного предприятия, описание аудитории скорее всего заменено рекламным охватом.

Соберите три объявления по одной интересующей вас теме и выпишите только проверяемые обещания. Уберите слова «умный», «новый», «уникальный» и «прорывной». Сравните оставшиеся действия. Часто выясняется, что два дорогих предложения не отличаются от дешёвого по тому, что вы действительно сможете получить. Иногда, наоборот, небольшая деталь — выгрузка данных или человеческая проверка — объясняет существенную разницу в цене.

Эта работа занимает меньше времени, чем исправление неудачной покупки. Она возвращает вам роль человека, который выбирает услугу. После этого проще разобраться в самой технологии: какие действия ей можно поручить и почему уверенный ответ ещё не подтверждает правильности.

Глава 2. Что скрывается за словом «нейросеть»

Для оценки покупки не нужно уметь обучать большую модель. Нужно различать модель, сервис и рабочую систему. Модель обрабатывает данные и формирует выход. Сервис добавляет интерфейс, тарифы, хранение и доступ. Рабочая система соединяет сервис с людьми, документами, правилами и действиями. Рекламная фраза нередко приписывает всей системе то, что показала только модель.

Языковая модель получает текст, который внутри представляется последовательностью небольших единиц — токенов. Токен может соответствовать слову, его части или знаку. Во время формирования ответа система рассчитывает продолжение с учётом доступного контекста. Такой механизм способен давать полезные объяснения и решать задачи, однако правдоподобное продолжение само по себе не является проверкой факта. Короткое описание этого риска есть в профиле генеративного ИИ NIST [1].

Одна из важных архитектур в развитии языковых моделей — Transformer, представленная в работе Vaswani и соавторов 2017 года. Авторы исследовали механизм внимания и задачи машинного перевода [2]. Из этой истории не следует рекламный вывод, что любая программа с похожей архитектурой обладает универсальной надёжностью. Архитектура объясняет устройство, а качество конкретного сервиса проверяют отдельно.

У покупателя есть полезная возможность: смотреть не на впечатление от разговора, а на путь данных. Вы вводите запрос. Сервис может дополнить его скрытыми инструкциями, найти сведения в интернете, обратиться к базе компании, вызвать калькулятор или другой инструмент. Затем он выдаёт ответ. Разные участки этого пути способны ошибаться по разным причинам.

Если ответ содержит неправильную сумму, проблема могла возникнуть при чтении документа, переносе цифр или вычислении. Если бот назвал старую цену, возможно, он нашёл устаревшую страницу. Если письмо ушло не тому человеку, ошибка могла быть в сопоставлении адресата. Слово «ИИ» не объясняет, где именно произошёл сбой. Продавец, который всё объясняет «капризами нейросети», мешает вам оценить пригодность услуги.

Контекст — информация, доступная системе при выполнении текущего задания. Длинный допустимый контекст означает, что сервис способен принять большой объём данных. Он не гарантирует, что каждая деталь будет правильно использована. Это примерно тот же вопрос, что и при передаче человеку большой папки: получение папки ещё не подтверждает внимательного чтения всех документов.

Обучение и подключение информации тоже не одно и то же. Продавец может сказать, что «обучил ИИ на вашей компании», хотя фактически загрузил инструкции и сделал поиск по файлам. Иногда этого вполне достаточно для задачи. Но тогда стоит обсуждать обновление базы, права доступа и точность поиска. Платить за обещанное дорогое обучение, когда вам предоставляют другую услугу, необязательно.

Различайте генерацию и извлечение. Генерация создаёт текст или изображение. Извлечение должно перенести существующую информацию: номер договора, сумму, дату, условие. В творческом задании дополнительный вымышленный элемент может быть полезным. В извлечении реквизитов он является ошибкой. Одну и ту же степень свободы нельзя оценивать одинаково в этих двух задачах.

Схожая разница существует между предложением действия и его исполнением. Чат может написать «я отправил письмо», хотя у него нет подключения к почте. Если подключение есть, оно всё равно может завершиться неудачей. Подтверждением служит запись о выполненной операции в нужной системе, а не формулировка модели. Этот принцип станет особенно важен в главе об агентах.

Не стоит и делать противоположную ошибку: считать модель бесполезной только потому, что в её основе статистические вычисления. Польза определяется результатом. Если она помогает быстрее подготовить качественный черновик, выгода реальна. Спор о том, «понимает» ли система текст как человек, мало помогает решить, стоит ли покупать конкретную подписку для конкретной работы.

Спросите продавца, какие внешние инструменты использует продукт. Есть ли поиск? Какие источники он просматривает? Можно ли увидеть документ, на который опирается ответ? Есть ли ограничения на действия? Что происходит при недоступности внешнего сервиса? Даже простой ответ на эти вопросы даёт больше информации, чем большой список технических сокращений.

Если разработчик не раскрывает все детали реализации, это ещё не свидетельствует о недобросовестности. У него могут быть коммерческие причины. Однако покупателю всё равно необходимы условия использования и проверяемое поведение. Для оценки автомобиля не требуется чертёж каждого узла, но нужны данные о характеристиках, обслуживании и возможность испытать его. С программой логика похожа.

Условный помощник для отдела продаж читает прайс, составляет предложение и создаёт черновик письма. Уже эти действия могут быть ценными. Но назвать его «продавцом, который заменяет сотрудника» можно только после изучения остальных задач: выяснения потребности, согласования условий, обработки возражений, ответственности за обещания и передачи заказа. Простое название не делает сложный процесс выполненным.

Составьте схему своей задачи словами: что входит, какие проверки нужны, что выходит и кто принимает решение. Если вы хотите сводку договора, входом будет договор, проверкой — сверка существенных условий, выходом — сводка со ссылками на пункты, решением — действие человека. Такая схема помогает увидеть, чего именно не хватает предлагаемой программе.

После этого слово «нейросеть» перестаёт быть самостоятельным аргументом. Оно становится описанием части инструмента. Следующий вопрос уже звучит точнее: на каком участке система может ошибиться и сколько будет стоить исправление?

Глава 3. Уверенный ответ и цена ошибки

Самая неприятная ошибка языкового сервиса выглядит как нормальный ответ. Текст связный, термины знакомые, тон уверенный. В нём может быть выдумана ссылка, перепутана дата или пропущено исключение. Читателю трудно заметить проблему, если он и обратился к программе потому, что не знает тему.

Поэтому проверка не должна зависеть от того, звучит ли ответ убедительно. Укажите, что считается подтверждением. Для цитаты это открытый источник и совпадение текста. Для расчёта — исходные числа и независимое вычисление. Для вывода о договоре — конкретный пункт и его окружение. Для утверждения о действии — запись в системе, где действие должно было произойти.

Фраза «ответь только правду» может выражать намерение пользователя, но не создаёт новую процедуру проверки. Точно так же просьба «не ошибайся» не исправляет неверную исходную таблицу. Если продавец обещает устранить все ошибки одним секретным запросом, спросите, как измерялся результат и какие случаи вошли в испытание.

Есть способы повысить опору ответа на документы. В работе Lewis и соавторов описан подход RAG: генерация соединяется с поиском внешней информации [3]. Для покупателя важен практический смысл: система может обращаться к материалам, а не полагаться только на сведения, усвоенные при обучении. Но наличие поиска не доказывает, что найден правильный документ или что ответ верно передаёт его содержание.

Условный бот отвечает на вопросы покупателей по инструкции магазина. В базе одновременно лежат старая и новая редакции правил возврата. Поиск находит подходящий по словам фрагмент старой версии. Ответ содержит ссылку, выглядит обоснованным и всё же неверен для текущей покупки. Проверять нужно версию источника и применимость условия, а не только наличие ссылки.

Другой условный пример: система видит пункт «скидка предоставляется при заказе от пятидесяти единиц» и сообщает, что скидка доступна всем корпоративным клиентам. Ссылка ведёт на существующий файл. Ошибка возникает уже при интерпретации. Поэтому полезно просить рядом с выводом показать короткий исходный фрагмент и затем сверять его самостоятельно.

Если ответа в базе нет, надёжное поведение иногда состоит в отказе. Для справочного помощника фраза «в материалах нет нужного условия» может быть лучше красивого предположения. Включайте такие случаи в проверку. Система, которая отвечает на всё без различия, производит впечатление компетентности, но это впечатление не показывает качество решения.

Документированный пример границы между помощником и специалистом — дело DoNotPay. FTC сообщила, что компания рекламировала сервис как роботизированного юриста и, согласно претензии регулятора, не провела достаточных испытаний соответствия уровню человеческого юриста. Итоговое распоряжение, одобренное 16 января 2025 года, ограничило неподтверждённые заявления такого рода и предусмотрело денежную выплату в 193 000 долларов [4]. Это американское регуляторное дело, а не решение российского суда и не доказательство бесполезности любых юридических инструментов.

Практический вывод из этого примера узок и полезен: чем серьёзнее заявленная замена специалиста, тем сильнее должны быть доказательства. Способность составить текст, похожий на юридический документ, не подтверждает пригодности документа для вашей ситуации. Между этими утверждениями находятся применимое право, факты, сроки, исключения и ответственность.

Цена ошибки зависит от задачи. Неподходящий заголовок можно удалить. Неправильную цену в отправленном коммерческом предложении придётся объяснять клиенту. Ошибка при

обработке чувствительных сведений может затронуть людей, которые вообще не участвовали в выборе программы. Поэтому одинаковая «точность 95 процентов» не означает одинаковой безопасности во всех процессах.

В условном наборе из ста заданий система допустила пять ошибок. Если все пять относятся к оформлению, ручная проверка может быть дешёвой. Если одна приводит к ошибочному платежу, общий процент успеха скрывает важную проблему. Отмечайте тяжесть ошибки отдельно от её количества. Иногда один редкий тип сбоя определяет, можно ли разрешать автоматическое действие.

Небольшое испытание даёт предварительную информацию, а не гарантию. Если на двадцати документах ошибок не было, нельзя заключить, что двадцать первый обязательно будет обработан правильно. Можно заключить только, что на выбранных двадцати документах при таких условиях ошибок не обнаружили. Эта формулировка кажется скромной, зато помогает принять разумное решение о следующем этапе.

Проверять всё вручную тоже не всегда рационально. Если инструмент используется для идей, достаточно выбрать пригодное и отвергнуть остальное. Если он переносит суммы из счетов, нужна более строгая сверка. Назначайте контроль по последствиям, а не по моде. Автоматизация должна сокращать суммарную работу, включая проверку и исправления.

Запишите для своей задачи три возможные ошибки и действие при каждой. Что произойдёт, если источник устарел? Если данных не хватает? Если система выполнила не то действие? Когда у вас есть ответы, обсуждение инструмента становится предметным. Следующая задача — убедиться, что продавец демонстрирует именно эти условия, а не тщательно выбранную удачу.

Глава 4. Что остаётся за кадром демонстрации

Демонстрация отвечает на вопрос, может ли продукт показать определённый результат. Покупателю нужен другой ответ: сможет ли он получать приемлемый результат на своих задачах с понятными расходами. Между этими вопросами находится работа, которую легко убрать из рекламного видео.

Условный продавец показывает, как за минуту создаёт изображение для карточки товара. В ролике не видно, откуда появилась исходная фотография, сколько вариантов было отвергнуто, сколько времени заняла правка и соответствует ли изображение настоящему товару. Возможно, инструмент действительно удобен. Однако минутная запись не устанавливает минутную себестоимость готовой карточки.

Попросите показать непрерывный путь: исходные материалы, ввод задания, ожидание, результат, исправление и выгрузку. Если видео сокращено, спросите, сколько времени исключено из записи. Монтаж не запрещён и часто необходим для понятного рассказа, но сокращённую демонстрацию нельзя использовать как точный замер труда.

Особенно важно число попыток. Один удачный результат из трёх и один из тридцати имеют разную стоимость. Даже если повторная генерация входит в подписку, время человека не исчезает. Если за каждую попытку списываются внутренние единицы оплаты, которые сервис называет кредитами, различается и прямой расход. Красивый финал без числа попыток показывает качество выбранного образца, но не эффективность процесса.

Проверьте, насколько задача похожа на вашу. В рекламной записи показывают документ с крупным шрифтом и без таблиц. У вас скан с пометками, печатью и двумя колонками. На сцене бот разговаривает с сотрудником разработчика, который произносит знакомые фразы. Ваши клиенты говорят с паузами, меняют тему и задают неожиданные вопросы. Разница входных данных может оказаться важнее различия моделей.

Есть и скрытая подготовка. Разработчик заранее очищает документы, исправляет названия файлов, добавляет правила и подбирает примеры. Такая работа может быть частью добросовестного внедрения. Проблема возникает, когда покупателю обещают тот же результат без подготовки и не включают её в цену. Спросите, что нужно привести в порядок до запуска и кто это делает.

Если на демонстрации специалист вмешивается, уточните его роль. Он выбирает лучший вариант? Исправляет ошибку? Подтверждает отправку? Подсказывает системе следующий шаг? Любое вмешательство допустимо, если оно соответствует обещанной услуге. Но демонстрация с постоянной помощью не доказывает автономной работы.

Условный сервис создаёт презентацию за пять минут. После этого дизайнер ещё сорок минут выравнивает графики, меняет шрифты и проверяет цифры. Результат может быть выгоднее прежнего двухчасового процесса. Честное сравнение будет «сорок пять минут вместо двух часов». Сравнение «пять минут вместо двух часов» исключает обязательную часть производства.

Попросите показать неудачный случай. Этот вопрос часто полезнее просьбы о ещё одном красивом образце. По реакции видно, обсуждает ли продавец эксплуатацию или поддерживает впечатление безупречности. Разработчик, знакомый со своим продуктом, обычно способен объяснить типичные ограничения. Если любой сбой приписывается только глупости пользователя, вам трудно будет получить помощь после покупки.

Неудачный случай нужен и для оценки восстановления. Система ошиблась; что дальше? Можно ли повторить только испорченный этап? Сохранились ли исходные данные? Есть ли журнал? Вернутся ли потраченные кредиты? Сколько действий выполнено до обнаружения проблемы? Продукт оценивают по поведению при сбое так же, как по успеху.

Слово «живое» не всегда означает независимое испытание. Продавец может заранее знать, какое задание будет выполнено, и подготовить систему. Предложите похожую задачу, которую не давали раньше. Сохраните разумность требований: если продукт предназначен для стандартных счетов, не нужно проверять его на рукописи XVIII века. Независимость испытания означает новый типичный вход, а не искусственно невозможное условие.

Присматривайтесь к завершению процесса. На экране появилась красивая картинка, но можно ли её скачать в нужном разрешении? Программа создала таблицу, но открывается ли файл и работают ли формулы? Бот сообщил об отправке, но появилось ли письмо в журнале? Покупателю нужен пригодный результат за пределами окна демонстрации.

Для оценки полезно взять одно утверждение и проверить его полностью. Например, «готовый ролик за десять минут». Заранее определите, что значит готовый: длительность, формат, отсутствие заметных дефектов, читаемые титры, допустимая музыка и возможность публикации. Затем засекайте время до этого состояния. Так слово «готовый» перестаёт быть украшением рекламы.

После демонстрации составьте две короткие записи. В первой укажите, что вы действительно увидели. Во второй — что осталось неизвестным. «Получен один пригодный образец» и «неизвестна доля брака на моих материалах» вполне могут стоять рядом. Не нужно ни отвергать продукт, ни заполнять пробелы доверием. Неизвестное проверяют отдельным испытанием.

Глава 5. Испытание, которое нельзя выиграть красивым ответом

Хороший тест начинается до знакомства с результатами. Если сначала увидеть ответ и только потом решить, что считать успехом, понравившийся продукт легко получит преимущество. Опишите задачу, условия и критерии заранее. Это защитит вас и от рекламы, и от собственной симпатии к удобному интерфейсу.

Возьмём условную задачу: извлечь из документов дату, сумму и номер договора. Успехом будет правильное заполнение всех трёх полей с указанием исходной страницы. Пропуск не считается правильным ответом. Выдуманное значение фиксируется отдельно. Если поле отсутствует, система должна обозначить отсутствие. Формат ответа тоже задаётся до испытания, иначе вы можете потратить время на преобразование результата.

Соберите небольшой набор обычных материалов, несколько сложных и несколько неполных. Обычные показывают повседневную полезность. Сложные выявляют границы. Неполные показывают, как система ведёт себя, когда решение невозможно. Не берите только те примеры, на которых продукт уже хорошо показал себя. И не используйте документы с чужими чувствительными сведениями без необходимого основания.

Для творческой работы критерии будут другими. Условный рекламный текст должен сохранить свойства товара, не добавить неподтверждённых обещаний, уложиться в заданный объём и соответствовать тону бренда. Оценка может включать человеческое суждение, но часть условий всё равно проверяется прямо. Наличие вкусовой оценки не оправдывает отсутствие критериев вообще.

Разделите обязательные условия и желательные качества. Если файл не открывается, красивый стиль уже не спасает результат. Если в объявлении неверная цена, высокая выразительность не делает его пригодным. Сначала проверяют обязательное. Только затем сравнивают удобство, качество формулировок и скорость.

Проведите сравнение с текущим способом работы. Если раньше задача занимала двадцать минут, а с ИИ — десять минут подготовки, пять минут ожидания и десять минут проверки, экономии нет. Если сотрудник может выполнять другую работу во время ожидания, учитывайте отдельно активное время и общее время выполнения. Для срочного ответа важна задержка, для себестоимости — занятость человека.

Условный тест включает тридцать документов. Полностью верно обработаны двадцать четыре, в четырёх требуется небольшая правка, в двух пропущены существенные данные. Можно сообщить «80 процентов документов обработаны без правок». Нельзя сообщить «почти всё автоматизировано», если ручная проверка каждого файла остаётся обязательной. Размер выборки и характер ошибок должны сопровождать цифру.

Записывайте затраты на исправление. Иногда продукт ошибается редко, но его ошибки трудно найти. Иногда он ошибается чаще, зато явно помечает сомнительные места. Для рабочего процесса второй вариант может быть удобнее. Сравнение только по доле правильных ответов пропускает стоимость контроля.

Не разрешайте продавцу менять условия посередине без фиксации. Он может предложить улучшить инструкцию или добавить документ — это нормально. Тогда проведите новый этап и отдельно сохраните старый результат. Иначе настройка будет продолжаться до первого успеха, а весь путь назовут первоначальной точностью продукта.

Версия и настройки входят в описание испытания. Сохраните дату, название тарифа, доступные функции, инструкцию и входные материалы. Если сервис обновится, вам будет с

чем сравнивать поведение. Скриншот ответа полезен, но файл с исходным заданием и результатом обычно сообщает больше.

Повторяемость тоже важна. Для нескольких типичных задач выполните повторный запуск. Если результаты заметно различаются, оцените, допустимо ли это. Для идей разнообразие может быть достоинством. Для переноса реквизитов оно требует дополнительного внимания. Не объявляйте различие автоматически дефектом; соотнесите его с назначением продукта.

Более строгий тест можно сделать слепым: убрать названия сервисов из результатов и попросить компетентного человека сравнить их по одинаковым критериям. Это особенно полезно для текста и дизайна, где известный бренд создаёт ожидание качества. Если оцениваете самостоятельно, хотя бы перемешайте ответы и не смотрите на цену до первой оценки.

Результат небольшого теста лучше формулировать как решение о следующем шаге. «Подходит для черновиков при проверке человеком». «Нужен дополнительный тест на скалах». «Не допускается к отправке писем». «Цена не оправданна при нашем объёме». Такие выводы полезнее общего балла, который не показывает способ применения.

Есть задачи, где самостоятельно установить правильность вы не можете. Тогда испытание должен оценить специалист или нужно выбрать менее ответственное применение. Просить вторую нейросеть проверить первую иногда помогает заметить расхождение, но совпадение двух ответов не является независимым доказательством истины. Они могут использовать один и тот же ошибочный источник.

В конце теста вернитесь к первоначальному обещанию. Если продавали экономию, проверьте суммарное время. Если продавали автономность, посчитайте вмешательства. Если продавали доход, технический тест ещё не завершает проверку: нужны покупатели и экономический расчёт. Чтобы оценить числовые заявления продавца, нужно также разобраться, как рассчитаны проценты точности.

Глава 6. Проценты точности и удобные знаменатели

Число «99 процентов» кажется точнее любого описания. Но процент имеет смысл только вместе с тем, что считали. Точность распознавания отдельных символов, правильность полей, пригодность документов и успех полного процесса — разные показатели. Одинаковая цифра может скрывать совершенно разную работу.

Условный документ содержит двадцать полей. В ста документах всего две тысячи полей. Система правильно перенесла 1 980, а двадцать ошибочных распределились по двадцати документам. Точность полей составила 99 процентов, доля документов без ошибок — 80 процентов. Оба числа правильны. Для человека, отправляющего документы дальше, второе может быть важнее первого.

Поэтому до сравнения попросите единицу оценки. Что считается одним заданием? Можно ли частично правильный ответ считать успехом? Ошибка в незначительном поле и в сумме учитывается одинаково? Если продавец не может определить показатель, цифра пока не пригодна для решения.

Есть проблема неравномерных случаев. Условная система выявляет редкие опасные сообщения. Из тысячи сообщений опасны десять. Программа всегда говорит «всё нормально» и правильно оценивает 990 обычных сообщений. Общая точность — 99 процентов, но ни одного опасного случая она не обнаружила. Для этой задачи показатель выглядит отличным и одновременно скрывает полный провал цели.

Нужно отдельно смотреть, сколько нужных случаев найдено и сколько обычных ошибочно помечено. Если система часто поднимает ложную тревогу, люди тратят время и перестают обращать внимание. Если пропускает важное, последствия могут быть серьёзными. Не существует одного процента, который без условий решает оба вопроса.

Сравнение с человеком тоже требует равных условий. Человеку дали грязный скан и ограничили время, программе — чистый файл и подготовленные инструкции. Или наоборот. Спросите, что получили обе стороны, сколько времени имели и кто оценивал результат. Без этого «превозмогли эксперта» может описывать устройство теста, а не продукт.

Бенчмарк — набор заданий и правил оценки для сравнения систем. Он полезен, когда соответствует вопросу. Но высокий результат на математических задачах не подтверждает качество разговора с клиентом. Хороший перевод не подтверждает точность бухгалтерского извлечения. Оценка относится к испытанному типу задач, а перенос требует обоснования.

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.