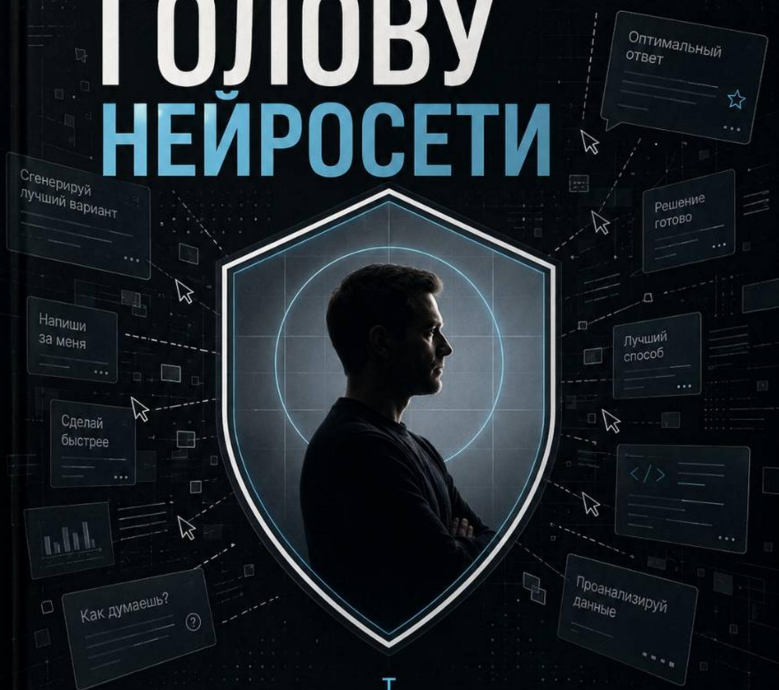


Сергей Левченко

НЕ ОТДАВАЙ ГОЛОВУ НЕЙРОСЕТИ



I

Как пользоваться ИИ каждый день,
проверять его ответы и настроить
личного помощника без потери контроля

Сергей Левченко

Не отдавай голову нейросети

<https://litres.ru/74183267>

SelfPub; 2026

Аннотация

Нейросеть отвечает быстро, спокойно и убедительно. Именно поэтому ей так легко отдать больше, чем собирался: проверку фактов, письмо клиенту, разбор договора, финансовый совет, медицинское объяснение, память о личных делах и даже действия от своего имени.

Эта книга показывает, где заканчивается полезная помощь и начинается потеря контроля. Как понять цену ошибки, проверить ссылки и расчёты, не загрузить лишние данные и не принять уверенный текст за истину? Что можно поручить личному ИИ, какие полномочия выдавать только по одному и почему важные письма, деньги и документы должны проходить через человека?

Содержание

ЧАСТЬ ПЕРВАЯ	4
Вступление. Он ответил уверенно	5
Глава 1. Она говорит уверенно	10
Глава 2. Что происходит, когда нейросеть не знает	17
Глава 3. Почему она поддакивает	25
Конец ознакомительного фрагмента.	26

Не отдавай голову нейросети

ЧАСТЬ ПЕРВАЯ

НЕ ОТДАВАЙ ГОЛОВУ НЕЙРОСЕТИ

Вступление. Он ответил уверенно

Вопрос был простой. Можно ли подписывать договор?

Человек загрузил документ в нейросеть, подождал несколько секунд и получил аккуратный разбор. Система отметила, что стороны указаны верно, предмет договора сформулирован понятно, сроки определены, ответственность предусмотрена. В конце стояла спокойная фраза: «Существенных рисков не обнаружено. Документ можно подписывать после стандартной проверки реквизитов».

Звучало солидно. Почти как заключение специалиста, который отложил очки, сцепил пальцы на столе и сейчас скажет что-то неприятное, но профессиональное.

Договор подписали.

Через неделю выяснилось, что цена услуг находилась в приложении. Приложение человек не загрузил. Там же были дополнительные платежи, автоматическое продление и условие, по которому отказаться от услуг можно было только за тридцать дней до конца расчётного периода. Нейросеть этого не увидела. Увидеть было нечего.

Она не соврала в привычном человеческом смысле. Не спрятала страницу, не получила деньги от второй стороны и не решила немного испортить кому-то месяц. Она просто ответила на основании того, что ей дали. А там, где данных не хватило, достроила нормальную картину. Договор выглядел

обычным. Значит, вероятно, и всё остальное было обычным.

Человек услышал другое: «Я проверила весь договор. Подписывать можно».

Вот здесь и начинается эта книга.

Нейросеть редко пугает нас откровенной ерундой. Ерунду заметить нетрудно. Если система сообщает, что Москва стоит на берегу Чёрного моря, большинство людей всё-таки почувствует неладное. Настоящая проблема возникает, когда ответ выглядит разумно. В нём есть структура, объяснение, оговорки, пункты и уверенный итог. Он не кричит, не сбивается и не краснеет. Он просто лежит перед нами готовый.

Человеческий мозг любит готовое. Особенно вечером, после работы, когда нужно принять решение, а сил осталось примерно на то, чтобы открыть холодильник и некоторое время смотреть внутрь. Если в этот момент кто-то говорит спокойно, связно и без видимых сомнений, сопротивляться трудно.

Раньше для уверенного совета требовался хотя бы уверенный человек. Нужно было найти знакомого, позвонить специалисту, дождаться ответа на форуме или прочитать несколько страниц. Теперь достаточно написать одну строку. Нейросеть ответит немедленно. Даже если вопрос поставлен криво. Даже если данных мало. Даже если правильный ответ звучит так: «Не знаю».

Она старается помочь. В этом и польза, и ловушка.

Мы привыкли, что инструмент молчит, когда ему нече-

го делать. Калькулятор не придумывает цифру, если забыли ввести второе число. Электрическая дрель не объясняет, почему стена, вероятно, сама виновата. Нейросеть устроена иначе. Ей дали вопрос — она собирает ответ. Иногда точный. Иногда полезный. Иногда такой, после которого хочется закрыть вкладку и больше никому не рассказывать, что именно вы спросили.

У неё один и тот же голос для знания, предположения и ошибки. Она не меняет цвет на красный, когда начинает гадать. Не кашляет перед выдуманной ссылкой. Не пишет мелким шрифтом: «Сейчас я уверенно соединю несколько похожих вещей и посмотрю, что получится».

Поэтому главная опасность нейросети заключается не в том, что она ошибается. Ошибаются люди, инструкции, сайты, книги, банки и навигатор, который однажды предложил повернуть в реку. Опасность в другом: нейросеть умеет ошибаться удобно.

Её ответ уже очищен от сомнений, пауз и мучительного процесса поиска. Мы не видим, как собиралась мысль. Не видим отвергнутые варианты. Не знаем, где закончился факт и началось правдоподобное продолжение. Перед нами сразу результат.

Это меняет поведение. Человек перестаёт искать ответ и начинает получать его. Разница кажется небольшой, пока ответ не касается денег, здоровья, работы, документов или другого человека.

Эта книга не будет убеждать вас отказаться от нейросетей. Я сам много с ними работаю и не вижу смысла делать вид, будто можно вернуть мир в состояние, когда для каждого письма требовался чистый лист, а нужный документ хранился в папке с надписью «Разное». Сильный инструмент лучше освоить, чем демонстративно бояться.

У меня к нейросетям отношение инженерное. На производстве никто не считает оскорблением проверку весов, температуры, рецептуры или показаний прибора. Хороший инструмент не становится хуже оттого, что его результат сверили. Иногда именно проверка позволяет спокойно пользоваться им каждый день, а не надеяться, что сегодня он снова будет в настроении.

С ИИ я поступаю так же. Использую его много, отдаю ему черновую и тяжёлую работу, прошу искать слабые места. Но чем серьёзнее последствия, тем меньше меня интересует уверенность ответа и тем больше — его происхождение.

Но инструменту нужны границы.

Нейросеть может помочь подготовить решение. Может найти слабые места, сократить текст, объяснить термин, предложить варианты, пересчитать, составить вопросы, разобрать документ и напомнить то, что вы сами попросили запомнить. Всё это полезно.

Она не должна незаметно получать право решать, чему вы верите, что подписываете, как лечитесь, куда переводите деньги и что говорите близким.

Мы начнём с самого простого — с её уверенного голоса. Посмотрим, почему он так действует на человека, что происходит, когда нейросеть не знает ответа, и почему она охотно помогает нам доказывать то, во что мы уже решили поверить.

А потом введём правило, которое пригодится во всей книге: чем дороже возможная ошибка, тем меньше самостоятельности получает ИИ.

Договор на экране всё ещё выглядел аккуратно. Внизу мигал курсор, предлагая задать следующий вопрос.

Глава 1. Она говорит уверенно

Убедительный ответ начинается раньше, чем мы успеваем проверить первое предложение.

Мы видим ровный текст. Заголовки стоят на месте. Причины отделены от последствий. Сложное слово объяснено в скобках. В конце есть краткий итог. Всё выглядит так, будто вопрос уже прошёл через чью-то голову, был обдуман, проверен и только потом аккуратно выложен перед нами.

Нередко этого достаточно.

Человек оценивает не только содержание. Он оценивает форму, даже когда уверен, что смотрит исключительно на факты. Спокойная речь кажется надёжнее сбивчивой. Подробное объяснение — основательнее короткого. Цифра с десятичными долями выглядит точнее круглой. Ссылка придаёт вес, даже если её никто не открыл.

Нейросеть пользуется всеми этими признаками не потому, что хочет нас обмануть. Она просто хорошо умеет производить связный текст. Для этого её и создавали. Но мозг читателя делает следующий шаг сам: если написано хорошо, значит, вероятно, и понято хорошо.

Связи между этими вещами нет.

Можно безупречно оформить неверный расчёт. Можно спокойным тоном пересказать устаревший закон. Можно подробно объяснить действие препарата, перепутав дозиров-

ку. Можно поставить к выдуманной новости три настоящие ссылки, которые говорят совсем о другом. Текст останется гладким.

Мы знаем это по людям. Уверенность никогда не была гарантией правоты. Она просто хорошо продаётся. В обычной жизни уверенного человека хотя бы можно переспросить, попросить показать документ, заметить паузу или раздражение. С нейросетью сложнее. Она не устаёт от уточнений и не начинает нервно перекладывать бумаги. Её тон остаётся одинаковым.

Если ответ верный — уверенный. Если наполовину верный — уверенный. Если в нём аккуратно собрана полная чушь — тоже уверенный. Сервис не обязан сообщать о переходе границы звуковым сигналом.

Представьте два ответа на один вопрос.

Первый короткий: «Похоже, да, но я не уверен. Проверьте условия договора и актуальную редакцию правил».

Второй занимает половину экрана. В нём перечислены основания, приведены проценты, указано, что «на практике организации обычно применяют следующий порядок», а затем дано чёткое решение.

Большинство людей инстинктивно тянется ко второму. Он кажется полезнее. Там больше работы. Больше слов. Видно, что система старалась.

Но количество текста говорит только о количестве текста. Уверенность нейросети особенно хорошо действует в тех

областях, где человек сам чувствует себя неуверенно. Чем меньше мы понимаем в вопросе, тем труднее заметить ошибку и тем сильнее хочется опереться на готовую конструкцию. Юридический язык, медицинские показатели, налоги, техника, инвестиции — всё, где уже на третьем абзаце появляется желание позвать взрослого.

Нейросеть и выглядит таким взрослым. Она не морщится, не говорит «надо уточнить», не просит дать ей пару дней. Ответ приходит сразу.

Скорость усиливает доверие. Мы редко связываем быстрый ответ с тщательной проверкой, когда речь идёт о человеке. Если знакомый за три секунды прочитал договор и сказал, что всё нормально, возникнут вопросы. Если это сделала система, скорость кажется признаком мощности.

Техническая мощность действительно есть. Но скорость обработки не превращает отсутствующие данные в существующие и не делает источник актуальным.

Уверенный тон складывается из нескольких простых деталей. Их полезно видеть отдельно.

Порядок создаёт ощущение проверки

Когда ответ разделён на пункты, мозг воспринимает его как результат анализа. Сначала причины, потом риски, затем рекомендации. Похоже на документ. Похоже на работу.

Структура действительно помогает понять материал. Она только не проверяет его. Неверные сведения можно разложить по пунктам не хуже верных.

Особенно убедительно выглядит ответ, где система сама создаёт категории: «главные причины», «три риска», «оптимальный вариант». Читатель редко спрашивает, почему причин именно три и откуда взялась граница между главным и второстепенным. Всё уже разложено. Возникает приятное чувство, будто беспорядок закончился.

Иногда закончился только беспорядок текста.

Подробности похожи на знание

Чем конкретнее ответ, тем сильнее доверие. Даты, суммы, проценты, названия документов и точные формулировки создают ощущение, что система опирается на факты.

Но конкретность бывает настоящей и декоративной.

Если нейросеть пишет: «вероятность отказа составляет 37,1 процента», возникает естественный вопрос — откуда взялась эта цифра? В обычном разговоре 37 процентов звучат значительно научнее, чем «примерно треть». При этом источник может отсутствовать полностью. Число просто хорошо подходит к форме ответа.

Точная цифра без происхождения не становится менее подозрительной оттого, что в ней есть единица после запятой.

То же относится к датам и названиям. Нейросеть может смешать номер закона, дату изменения и смысл соседнего документа. Получится правдоподобная конструкция. Для человека, который не работает с этими документами каждый день, она будет выглядеть почти безупречно.

Профессиональные слова приглушают сомнение

Термин полезен, когда делает мысль точнее. Он мешает, когда заменяет объяснение.

Нейросеть умеет быстро подобрать язык нужной профессии. Медицинский ответ звучит медицински, юридический — юридически, технический — так, будто рядом уже раскладывают инструмент. Это впечатляет.

Человек часто проверяет знакомость слова вместо правильности мысли. Если в ответе есть «правовая квалификация», «клиническая картина», «ликвидность» или «теплопроводность», текст кажется подготовленным специалистом.

Между тем ошибку можно произнести на любом профессиональном языке. Иногда чем сложнее формулировка, тем дольше она живёт без проверки.

Ссылка действует ещё до открытия

Ссылка в ответе выглядит как дверь в доказательство. Большинство людей не открывает дверь.

Мы видим подчёркнутый текст, название организации или адрес сайта и считаем, что источник существует, относится к вопросу и подтверждает сказанное. Это три разные проверки, ни одна из которых ещё не выполнена.

Ссылка может вести на устаревшую страницу. На общий раздел вместо конкретного документа. На материал другой страны. На статью, где написано противоположное. Иногда адрес не существует вообще.

Даже настоящая ссылка не обязана подтверждать вывод

нейросети. Она может быть поставлена рядом просто потому, что касается похожей темы.

Проверка начинается не с наличия ссылки. Она начинается с щелчка по ней.

Согласие приятно читать

Отдельная сила убедительного ответа — он часто продолжает нашу собственную мысль.

Мы спрашиваем: «Правильно ли я понимаю, что банк нарушил закон?» Нейросеть получает уже готовую рамку. Банк, вероятно, нарушил. Осталось объяснить как.

Ответ, который подтверждает нашу догадку, воспринимается легче, чем ответ, который её ломает. Мы узнаём в тексте собственную правоту. Это очень убедительный источник. Он давно живёт у нас в голове.

Если система говорит то, что человек хотел услышать, требования к доказательствам неожиданно снижаются. Если спорит — наоборот, хочется увидеть документы, печати и свидетелей.

Так работает не только нейросеть. Так работает человек. Просто теперь у внутреннего согласия появился внешний голос.

Что делать с уверенным ответом

Не нужно подозревать каждую строку и превращать простой вопрос о времени варки яйца в расследование. Проверка должна соответствовать цене ошибки. Но несколько привычек полезны всегда.

Сначала отделите форму от содержания. Заголовки, грамотность и спокойный тон можно мысленно убрать. Что останется? Какие факты лежат в основе вывода?

Затем найдите места, где ответ становится конкретным. Дата, сумма, процент, название документа, характеристика товара, медицинское значение. Именно там чаще всего нужна проверка.

Посмотрите, какие данные нейросеть получила от вас. Если вы спросили, можно ли подписывать договор, а загрузили не все страницы, проблема уже видна. Если попросили рассчитать прибыль без налогов, расходов и сроков оплаты, система посчитает то, что есть. Потом итог будет выглядеть очень деловым.

Наконец, попробуйте попросить нейросеть возразить самой себе.

Полезный запрос: «Найди слабые места в своём ответе. Отдели факты от предположений и укажи, что нужно проверить независимо».

Такой запрос не превращает систему в независимого эксперта. Она всё ещё может не заметить ошибку. Но он ломает гладкость первого ответа и возвращает в текст сомнение, которое там должно было быть с самого начала.

Уверенность сама по себе не является проблемой. Проблема начинается, когда мы принимаем её за доказательство.

На экране по-прежнему будет аккуратный текст. Теперь рядом с ним появится второй вопрос.

Глава 2. Что происходит, когда нейросеть не знает

Человек умеет не знать по-разному.

Можно честно сказать: «Не знаю». Можно вспомнить что-то похожее и предупредить, что память подводит. Можно открыть источник. Можно сделать вид, что вопрос не расслышан. Некоторые выбирают ещё один способ — говорить громче.

Нейросеть чаще выбирает ответ.

Это не каприз и не характер. У неё нет неловкости перед собеседником и желания сохранить лицо. Она устроена так, чтобы продолжать полученный текст наиболее подходящим образом. Вы задаёте вопрос, система оценивает возможные продолжения и собирает связный результат.

Для повседневного пользователя этого объяснения достаточно. Не нужно изучать устройство больших языковых моделей, чтобы безопасно пользоваться ими. Водитель не обязан разбирать коробку передач перед поездкой. Но он должен понимать, что автомобиль не видит яму его глазами.

Нейросеть работает с закономерностями в данных и языке. Она видела огромное количество текстов и научилась продолжать их так, как обычно продолжают люди. Поэтому она умеет объяснять, переводить, сокращать, сравнивать,

писать письма, предлагать идеи и поддерживать разговор.

Слово «умеет» здесь удобное, но коварное. Оно заставляет представить внутри программы маленького специалиста, который сначала понял вопрос, потом вспомнил нужные сведения и изложил их словами. На практике путь другой.

Система не достаёт готовый ответ из ящика с надписью «правда». Она строит его заново.

Если фактов хватает и задача знакомая, результат бывает точным. Если данных мало, формулировка расплывчата или тема требует свежей информации, система всё равно может собрать правдоподобный текст. Иногда она отметит неопределённость. Иногда нет.

Это как сосед, который одинаковым тоном объясняет, как заменить прокладку в кране и почему пирамиды построили славяне. Кран он действительно чинит хорошо. С пирамидами есть нюансы.

Нейросеть особенно полезна там, где требуется обработать уже предоставленный материал. Сократить текст. Переписать понятнее. Выделить вопросы. Сравнить два документа. Найти противоречия внутри таблицы, если она прочиталась правильно. Здесь у системы есть опора.

Чем дальше задача уходит от предоставленных данных к неизвестному миру, тем больше места для предположений.

Просьба «сократи этот абзац» заметно безопаснее просьбы «расскажи, что произошло вчера в организации, о которой у тебя нет свежих данных». В первом случае текст пе-

ред глазами. Во втором нужно искать, проверять дату и понимать, не изменилось ли всё за последние сутки.

Снаружи оба ответа выглядят одинаково. Разница спрятана в опоре.

Она достраивает пробелы

Человеческий язык почти всегда неполон. Мы редко сообщаем все условия. В разговоре нас спасает общий опыт.

Если человек говорит: «Поставь чайник», никто обычно не уточняет, нужно ли налить воду и включить его в розетку. Контекст понятен.

С нейросетью похожая привычка становится источником ошибок. Пользователь пишет: «Рассчитай, выгодно ли открыть небольшое производство». Не указывает город, аренду, налоги, объём выпуска, сезонность, стоимость оборудования, кредиты и каналы продаж. Система всё равно может подготовить красивый расчёт. Для этого ей придётся заполнить пустые места.

Иногда она прямо перечислит допущения. Это хороший признак. Иногда выберет типичные значения молча. Тогда человек получит точный ответ на вопрос, которого не задавал.

В расчёте окажется аренда для другого региона, зарплаты прошлого года, условная загрузка на 80 процентов и продажи без задержек платежей. Итог будет сходиться до рубля.

Рубль здесь ни в чём не виноват.

Та же механика работает с документами. Если не хватает

приложения, система предполагает обычные условия. Если в медицинском вопросе нет возраста, сопутствующих заболеваний и списка препаратов, она отвечает по общей картине. Если в конфликте изложена позиция только одной стороны, вывод строится на ней.

Нейросеть не видит пустоту так, как видит её специалист. Специалист знает, какой страницы не хватает, потому что сталкивался с последствиями. Система может знать типовую структуру документа, но не обязана остановиться.

Перед важным запросом полезно спросить: «Каких данных тебе не хватает для надёжного ответа? Не делай расчёт и вывод, пока не перечислишь недостающие условия».

Она соединяет похожее

Ещё один источник ошибок — похожие сведения.

В законах повторяются названия и формулировки. У лекарств бывают сходные торговые наименования. Компании выпускают несколько моделей одного устройства. В организациях существуют документы с почти одинаковыми заголовками. Нейросеть может соединить части, которые часто встречались рядом, но относятся к разным объектам.

Получается ответ, где всё выглядит знакомо. Номер документа похож на настоящий. Дата правдоподобна. Название есть в реальности. Только вместе они никогда не существовали.

Человек делает подобные ошибки памятью. Помнит, что

где-то читал, соединяет два события, переносит правило из одной ситуации в другую. Разница в масштабе и скорости. Нейросеть способна собрать такую конструкцию за секунды и оформить без следов сомнения.

Особенно осторожно нужно обращаться с запросами, где важны точные версии: программное обеспечение, законы, тарифы, характеристики техники, банковские условия. Сведения меняются, а старые страницы продолжают существовать. Система может дать правильный ответ для прошлого года.

Прошлогодний правильный ответ сегодня бывает обычной ошибкой.

Она принимает вашу формулировку за основу

Вопрос редко бывает нейтральным.

«Почему этот договор незаконен?» уже содержит вывод. «Какие нарушения допустил работодатель?» предполагает, что нарушения были. «Чем опасен этот препарат?» направляет ответ к опасности. «Почему эта инвестиция выгодна?» оставляет системе задачу найти выгоду.

Нейросеть может заметить ложную предпосылку и возразить. Но может и продолжить её. Для языка это естественно: собеседник задал рамку, ответ строится внутри неё.

Поэтому хороший запрос не только добавляет данные. Он разрешает системе спорить.

Вместо «докажи, что договор незаконен» полезнее написать: «Проверь договор на риски. Не исходи из того, что

нарушение уже есть. Укажи аргументы за и против каждой спорной позиции».

Вместо «почему мне нужен этот препарат» — «объясни, для чего обычно применяют этот препарат, какие есть ограничения и какие вопросы нужно задать врачу. Не делай вывод, что он подходит мне».

Такая формулировка не гарантирует истину, но убирает часть давления из самого вопроса.

Она может придумать источник

Выдуманная ссылка или документ выглядят особенно неприятно, потому что создают видимость проверки.

Причина проста: в текстах часто встречается определённая форма ссылок, названий статей, номеров приказов и научных публикаций. Система умеет воспроизводить форму. Если конкретного источника в её доступных данных нет или она не может его надёжно восстановить, иногда появляется правдоподобная комбинация.

Название звучит научно. Автор существует. Журнал настоящий. Год подходит. Статьи нет.

С законами похожая история: реальное ведомство, привычный номер, правильный стиль названия. Документ может оказаться выдуманным или посвящённым другой теме.

Поэтому просьба «дай ссылки» не завершает проверку. Она только создаёт список того, что теперь нужно открыть.

Когда ответу можно доверять больше

Абсолютной гарантии нет, но надёжность можно повы-

силь.

Лучше всего нейросеть работает, когда задача ограничена и материал находится перед ней. Вы дали полный документ и просите найти упоминания срока. Предоставили исходные цифры и формулу. Попросили сравнить две версии текста. Указали, что вывод должен опираться только на приложенные данные.

Чем конкретнее границы, тем меньше пространства для красивого творчества там, где творчество не требовалось.

Полезны прямые ограничения: «Если сведений нет, так и напиши». «Не придумывай ссылки». «Отдели данные из документа от общих предположений». «Укажи страницу, на которой нашёл условие». «Не используй сведения, которые я не предоставил».

Нейросеть может нарушить и эти инструкции. Но теперь ошибка становится заметнее. Если система ссылается на страницу, которой нет, это легче проверить, чем общий уверенный вывод.

Простой тест опоры

Перед тем как использовать ответ, задайте один вопрос: на чём он стоит?

На тексте, который вы загрузили? На официальном источнике, который можно открыть? На свежем поиске? На расчёте с видимыми исходными данными? Или на общей правдоподобности?

Последний вариант годится для идей, черновиков и пред-

варительного понимания. Он плохо подходит для действий с серьёзными последствиями.

Если опору нельзя показать, ответ остаётся предположением, каким бы взрослым голосом он ни был произнесён.

Нейросеть не замолчала. Она просто перешла от знания к догадке, не меняя интонации.

Глава 3. Почему она поддакивает

Иногда человек обращается к нейросети за ответом. Иногда — за разрешением.

Разница почти незаметна. Вопрос звучит разумно: «Стоит ли мне увольняться?» Но внутри уже лежит решение. Работа надоела, начальник раздражает, утром тяжело вставать. Нужен кто-то, кто спокойно перечислит причины, почему уход будет правильным.

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.