

12+

ИИ ДЛЯ СОМНЕВАЮЩИХСЯ

ИИ БЕЗ СТРАХА

Что спросить,
как проверить,
чему не верить



СПРАВОЧНИК ДЛЯ ТЕХ, КТО УЖЕ ПРОБОВАЛ И БРОСИЛ

Ден Ах ИИ-интегратор, более 50 проектов

Ден Ах

**ИИ без страха. Что спросить,
как проверить, чему не верить**

«Издательские решения»

Ден Ах

ИИ без страха. Что спросить, как проверить, чему не верить / Ден Ах — «Издательские решения»,

Вы уже пробовали ИИ — и поняли, что он иногда врёт? Тогда эта книга для вас. «ИИ без страха» — это не лекция о том, «как стать гением». Это практический справочник, который учит:

- Что реально спросить у нейросети
- Как быстро проверить любой ответ
- Чему на самом деле нельзя верить

Автор — Ден Ах, ИИ-интегратор многолетним опытом. Книга написана простым языком. Здесь только самое важное: вопросы + проверка + то, чему не верят. Никакой воды. Только рабочий инструмент.

© Ден Ах

© Издательские решения

Содержание

Введение. Эта книга не про кнопки	6
Как читать	7
Откуда я это знаю	8
Глава 1. Почему вы бросили в первый раз	9
Насмотренность вместо правил	15
Глава 4. Двенадцать слов, которые нужно знать	22
Что с этим делать	28
Глава 6. Кто и как ей уже пользуется	30
Чем именно занимаются	32
Конец ознакомительного фрагмента.	34

ИИ без страха Что спросить, как проверить, чему не верить

Ден Ах

© Ден Ах, 2026

ISBN 978-5-0071-3527-6

Создано в интеллектуальной издательской системе Ridero

Введение. Эта книга не про кнопки

В книге 2023 года про нейросети была подробная глава о том, как получить картинку: зайти в мессенджер, найти канал, набрать команду с косой чертой и параметрами через дефисы. Через год у сервиса появился обычный сайт с кнопками, и глава умерла.

Я держал эту историю в голове всё время, пока писал. Инструкции к экранам устаревают за месяцы, а книга живёт годами. Поэтому здесь минимум кнопок.

Здесь про другое: как эта штука устроена, почему она врёт, где ей верить и где проверять. И что с ней делать обычному человеку — у которого нет ни времени на курсы, ни желания становиться специалистом.

Для кого это написано

Для человека, который уже пробовал — и бросил.

Не для новичка, который ничего не знает. Для того, кто в обеденный перерыв открыл чат, задал вопрос, получил гладкий текст ни о чём и закрыл вкладку с мыслью «ну и зачем это всё».

Таких много, и это измерено — хотя с цифрой всё непросто, и я сразу покажу, как именно.

Устойчиво измеряется другое: доля разочарованных растёт. По опросу SuperJob двадцать четыре процента россиян считают, что искусственный интеллект бесполезен для их задач. Годом раньше так думали восемнадцать. Прирост пользователей идёт, а вместе с ним идёт и прирост тех, кто попробовал и не нашёл смысла.

Это говорит не о людях, а о том, как инструмент показывают.

Если вы никогда не пробовали — книга тоже подойдёт, просто первая глава будет для вас не про прошлое, а про будущее: там разобраны три места, где обычно обрывается первая попытка.

Что вы узнаете про себя

Обещаю не пользу вообще, а конкретное.

Вы поймёте, что ваша неудача была закономерной, а не следствием неспособности: инструмент устроен не так, как поисковая строка, и об этом никто не предупреждает.

Вы научитесь отличать место, где этой программе можно верить, от места, где нельзя. Это главный навык всей книги, и он важнее любых приёмов: без него человек либо не пользуется вообще, либо однажды приносит в важный документ то, чего не существует.

Вы перестанете тратить время на приёмы, которые не работают. Половина того, что советуют в статьях и продают на курсах, проверялась в лабораториях — и не подтвердилась. Целая глава посвящена тому, что можно спокойно выбросить.

И вы получите несколько привычек, которые останутся с вами независимо от того, какие сервисы будут через три года. Их всего пять, и в заключении я называю их прямо.

Чего в этой книге нет

Обещания стоят дёшево, поэтому сразу назову границы.

Нет устройства моделей изнутри. Ни трансформеров, ни слоёв, ни математики. Вам не нужно знать, как устроен катализатор, чтобы водить машину.

Нет программирования. Ни строчки кода.

Нет инструкций по обходу ограничений и оплате зарубежных сервисов. Тема существует, я её не обхожу молчанием и объясняю, как обстоят дела. Но подробных руководств здесь не будет: это юридически рискованно и для меня, и для площадки, где книга продаётся.

Нет обещаний, что нейросеть заменит вам врача, юриста и бухгалтера. Наоборот: есть отдельная глава про то, где проходит стоп-линия, с задокументированными случаями, чем кончается её переход.

И нет восторга. Тема им перегрета, а вы, скорее всего, на него уже один раз купились.

Как читать

Книга разбита на пять частей, и они идут от устройства к жизни.

Первая часть — что это такое и почему у вас не получилось. Вторая — с чего начать сегодня: сервисы, деньги, первый запрос, приёмы. Третья — самая важная: как проверять. Четвёртая — про жизнь: дом, работа, деньги, здоровье, дети, мошенники. Пятая — про страхи и про то, как не отстать.

Главы устроены так, чтобы читать их можно было по одной и вразнобой: каждая заканчивается коротким итогом и одним действием на сегодня. Если вы человек занятой, читайте по главе в день — за месяц пройдёте всю книгу и попробуете двадцать шесть разных вещей.

Если хотите быстрый результат, начните с трёх глав: первой, восьмой и тринадцатой. Первая объяснит, что случилось в прошлый раз. Восьмая покажет, как написать запрос, чтобы результат был. Тринадцатая — как проверить, что вам не соврали. Этих трёх хватит, чтобы инструмент начал приносить пользу уже сегодня вечером.

Ещё одна особенность, о которой стоит предупредить. В книге много цифр — примерно по восемь-десять на главу, — и у каждой назван источник и год. Это может показаться избыточным для книги с пометкой «для сомневающихся». Я делаю так сознательно, по двум причинам.

Первая: тема забита утверждениями без источника. «Нейросети повышают продуктивность на сорок процентов», «ИИ заменит треть профессий», «искусственный интеллект уже умнее человека» — всё это вы слышали, и почти нигде не сказано, кто, когда и на ком это мерил. Я не хочу добавлять к этому шуму.

Вторая: у каждой цифры есть срок годности. Названный год позволяет вам понять, когда её пора перепроверить, — а в последней главе я прямо говорю, что в этой книге устареет первым.

Если цифры вам не нужны, читайте мимо: все выводы сформулированы словами, и ни один не требует запоминания процентов.

И последнее. В книге есть сквозной персонаж — Ирина, сорок семь лет, бухгалтер. Она придумана, и я говорю об этом каждый раз, когда она появляется. Её обеденный перерыв, с которого начинается первая глава, не придуман: так выглядит первая попытка у большинства.

С него и начнём.

Откуда я это знаю

Я работаю с языковыми моделями с 2023 года — с первых русскоязычных версий, когда они ещё путались в падежах. Интегрирую ИИ решения в бизнес процессы, пользуюсь нейросетями в жизни, а последние полтора года делаю музыку: треки, собранные вместе с нейросетями, выложены на стриминговые сервисы под тем же именем, что стоит на обложке. Захотите послушать и составить своё мнение — найти несложно.

Музыка, кстати, оказалась лучшей школой из всех. В тексте гладкую пустоту можно не заметить, в звуке она слышна сразу. Именно там я окончательно понял, что этот инструмент делает быстро то, что вы сами способны оценить, и подводит везде, где оценить не можете.

Но опыт — не доказательство. Поэтому вся фактическая часть книги стоит на опубликованных исследованиях и официальных документах: опросах ВЦИОМ и Росстата, работах университетских лабораторий, документации сервисов, решениях судов. У каждой цифры в тексте назван источник и год. Где источники противоречат друг другу — я показываю оба и говорю, что не знаю, где правда. Где у меня только личное наблюдение — прямо пишу, что это наблюдение, а не измерение.

Список источников целиком — в приложении.

Чем эта книга отличается от других

На русском языке про нейросети сейчас издано немало, и перед тем как сесть за свою, я разобрал четырнадцать книг — от переводных руководств до самиздата. Ниша устроена странно: она расколота надвое, и между половинами пусто.

С одной стороны — сборники готовых запросов: триста тридцать три промпта, пятьсот промптов, тысяча. Они хороши для человека, который уже понимает, что делает, и просто ищет шаблон. Новичку они дают ложное ощущение, что где-то есть волшебная формула, которую надо раздобыть.

С другой — вводные книги про устройство искусственного интеллекта, с примерами кода и объяснением, как обучаются модели. Хорошие книги, но написаны для тех, кто собирается в профессию.

Между ними — взрослый человек не из ИТ, которому эта штука нужна не для карьеры, а для жизни, и который боится не сложности, а того, что его обманут или что он что-то испортит. Ему не адресована ни одна из четырнадцати.

Есть и пять вопросов, на которые не отвечает ни одна из них. Как проверить ответ. Что нельзя загружать в чат. Как выбрать сервис под свою задачу. Как пользоваться этим в непрофессиональной жизни — с документами, здоровьем, детьми, пожилыми родителями. И что делать читателю, который начинает с нуля.

Эта книга — про эти пять вопросов. Отсюда и её устройство: треть объёма отдана проверке и границам, а не приёмам.

Глава 1. Почему вы бросили в первый раз

Ирина открыла чат с нейросетью в обеденный перерыв. Написала: «Составь отчёт по продажам». Через пять секунд получила текст про важность анализа динамики и необходимость учитывать сезонность. Ни одной её цифры. Ни одного её месяца. Прочитала, закрыла вкладку, вернулась к таблице в экселе. На этом её знакомство с искусственным интеллектом закончилось — на четыре месяца, до тех пор пока дочь не показала, как это делают люди помоложе.

Ирина — учебный персонаж этой книги, ей сорок семь, она бухгалтер. Я буду возвращаться к ней в каждой части, и каждый раз буду говорить, что она придумана. Но её обеденный перерыв не придуман. Так выглядит первая попытка у большинства.

Вас таких много, и это измерено

Летом 2026 года ВЦИОМ опросил три тысячи двести человек и выяснил, что семьдесят восемь процентов российских интернет-пользователей за последний год хоть раз обращались к нейросетям. Цифра звучит победно.

А теперь то место, где мне придётся быть аккуратным, потому что цифры расходятся.

В волне осени 2025 года ВЦИОМ спрашивал не только «пробовали ли», но и «как часто». Двадцать восемь процентов пользователей ответили тогда — только пробовал. В волне июня 2026 года на похожий вопрос отвечают иначе: пробовали и бросили — шесть процентов, годом раньше было десять.

Двадцать восемь и шесть. Разница слишком велика, чтобы объяснить её ростом опыта за девять месяцев. Скорее всего, дело в формулировках: в одной волне «только пробовал» — это ответ про частоту обращений, в другой «попробовал и бросил» — отдельная категория рядом с постоянными пользователями. Я не знаю точно, где правда, и не буду делать вид, что знаю. Честная формулировка звучит так: доля тех, кто зашёл и не вернулся, лежит где-то между шестью и двадцатью восемью процентами, и обе цифры опубликованы одной и той же службой.

Зато другое измеряется устойчиво. По опросу SuperJob 2026 года двадцать четыре процента россиян считают, что ИИ бесполезен для их задач. Годом раньше так думали восемнадцать. Среди тех, кому за сорок пять, — двадцать восемь процентов.

Вот это и есть твёрдый факт: пользователей становится больше, и одновременно растёт доля разочарованных. Не «каждый четвёртый ушёл», а «людей, которые попробовали и не нашли смысла, с каждым годом больше».

Обычно эти цифры подаются как отставание: вот, люди не разобрались, надо просвещать. Я читаю их иначе. Инструмент, которым один раз попробовали воспользоваться четверо из пяти, а регулярно пользуется меньше половины, — это не проблема людей. Это проблема того, как инструмент показали.

И есть третья цифра, самая говорящая. Сервис Chad, который считает статистику по своим пользователям, в 2025 году опубликовал среднюю длительность сессии: семь минут тринадцать секунд. Люди заходят на семь минут. За семь минут можно задать три вопроса, получить три общих ответа и решить, что вам продали пустышку.

Три места, где обрывается первая попытка

За три года я посмотрел на чужие первые попытки — и на свои. Провалы удивительно однообразны, их всего три.

Первое место — регистрация. Человек слышал про ChatGPT, потому что про него говорят в новостях. Он идёт туда. Дальше по обстоятельствам: не принимается российский номер, отклоняется почта на привычном домене, не приходит письмо с подтверждением, бесконечно крутится проверка «вы не робот». В разборах на тематических площадках за декабрь 2025 года перечислены все эти отказы разом. И они не случайность. Россия и Беларусь отсут-

ствуют в списке стран, где OpenAI обслуживает пользователей. В справке компании прямо сказано: вход из неподдерживаемой страны может привести к блокировке аккаунта.

Дальше происходит самое печальное. Человек делает вывод не «этот конкретный сервис мне недоступен», а «мне туда нельзя». И не идёт искать российские, которые работают без всяких сложностей и бесплатно. Про них он просто не знает — про них не говорят в новостях.

Второе место — пустое окно. Допустим, регистрация прошла. Открывается белое поле и мигающий курсор. И человек не знает, что туда писать.

Это описано прямым текстом в руководствах для начинающих на русском языке за 2025 год, и я слышал это от каждого второго, кому показывал. Не «я не понимаю, как работает» — с этим как раз проще. А «я не понимаю, что спрашивать». Поисковая строка приучила нас к двум словам. Здесь двух слов мало, а насколько мало — никто не объясняет.

Третье место — первый ответ. Человек всё-таки пишет. Пишет коротко, как в поиск: «составь отчёт по продажам», «напиши письмо», «придумай название». И получает ровно то, что просил: отчёт вообще, письмо вообще, название вообще.

Гладкое, вежливое, ни о чём. В обсуждениях такой текст называют водянистым, а его главный признак описывают точно: много слов при скудном содержании. И вот тут человек делает вывод, который стоит ему следующего года: «оно не умеет».

Оно умеет. Ему не сказали.

Почему это не ваша вина

Теперь важное. Всё, что я описал, — не следствие вашей неопытности. Это следствие устройства инструмента, о котором вас никто не предупредил.

Нейросеть отвечает на текст, а не на замысел. У вас в голове ваш отчёт: три филиала, падение в мае, сезонная история с оптовиками. У неё — четыре слова. Из четырёх слов ваш отчёт не выводится, и она делает то единственное, что может: собирает средний отчёт из всех, что видела. Он выходит бесполезным не потому, что модель слабая, а потому что она честно ответила на заданный вопрос.

При этом она не переспросит. Человек на её месте переспросил бы: какие продажи, за какой период, для кого отчёт. Модель почти никогда так не делает. Она отвечает всегда — даже когда спрашивать было бы разумнее.

Дальше срабатывает вторая ловушка. Ответ выглядит уверенно. У него нет интонации сомнения, нет «мне кажется» и «возможно, я не всё понял». Он звучит как готовый результат. Поэтому вы оцениваете его как результат — и видите, что результат никакой.

Из этого делается вывод об инструменте. А надо бы — о разговоре.

Есть, кстати, цифра, которая ставит всё на место. Осенью 2025 года OpenAI совместно с Национальным бюро экономических исследований разобрала около миллиона ста тысяч обезличенных сообщений пользователей — без чтения текстов людьми, через классификацию. Оказалось, что программирование занимает четыре процента разговоров. Четыре. А семьдесят три процента сообщений вообще не связаны с работой: люди спрашивают про быт, про здоровье, про документы, про то, как объяснить ребёнку задачу.

То есть это давно не инструмент для айтишников. Это инструмент для тех, кто спрашивает. Просто спрашивать надо иначе, чем в поиске.

Что говорят те, кто не начал

В том же опросе ВЦИОМ 2026 года у непользователей спросили, почему. Ответы стоит прочитать медленно.

Сорок четыре процента: нет подходящих задач. Тридцать шесть: не знаю, какие у этого возможности. Двадцать четыре: сомневаюсь в безопасности. Двадцать два: не знаю, как пользоваться.

Первый ответ — самый частый и самый обманчивый. «Нет подходящих задач» почти никогда не означает, что задач нет. Оно означает: я не знаю, какие мои задачи для этого годятся.

Человек, который весь вечер разбирался с претензией в интернет-магазин, не считает это задачей для нейросети. Он считает это жизнью.

Второй и четвёртый ответы вместе дают пятьдесят восемь процентов людей, которым просто никто не показал. Не научил — показал.

А третий ответ, про безопасность, я оставляю без спора. Он обоснован. Про то, что действительно не стоит отдавать чужому серверу, будет отдельная глава, и она получилась одной из самых длинных в книге.

Вам ещё и показывали не то

Есть четвёртая причина, которую в опросы не заносят.

Допустим, человек не сдался и полез искать, как этим пользоваться. Что он найдёт? Списки промптов. Триста тридцать три готовых запроса для бизнеса и маркетинга, пятьсот запросов для жизни и бизнеса, тысяча запросов ещё для чего-нибудь. Это реальные книги, они лежат в магазинах, и по данным «Читай-города» за 2026 год продажи книг об искусственном интеллекте выросли на сорок процентов за год — в основном как раз за счёт таких сборников.

Сборник промптов — вещь неплохая, если вы уже понимаете, что делаете. Если не понимаете, он даёт ложное ощущение: будто существует правильная формула, и надо её раздобыть. Человек копирует чужой запрос, вставляет, получает чужой результат — потому что запрос был написан под чужую задачу. И укрепляется в мысли, что дело в секретном знании, которого у него нет.

Секретного знания нет. Есть три-четыре приёма, подтверждённых измерениями, — и они все про то, что вы должны сообщить модели о своей задаче. А ещё есть пяток приёмов, которые все повторяют друг за другом и которые на моделях 2026 года не дают ничего или вредят. Им посвящена отдельная глава, самая злая в книге.

Вторая беда обучающих материалов — тон. «Это же элементарно», «просто напишите промпт», «любой школьник разберётся». Человек, у которого один раз не вышло, читает это как приговор себе и закрывает страницу.

Между тем ровно эта аудитория сейчас и приходит. Сервис Chad сообщил, что за 2025 год число его пользователей старше пятидесяти пяти выросло в два с половиной раза. Там же интересная деталь: эта группа, выбрав один сервис, почти не пробует другие — альтернативы смотрят девять процентов. Люди осваивают один инструмент и держатся за него, потому что второй заход стоит им слишком дорого.

Татьяна, пятьдесят восемь лет, преподаватель — второй учебный персонаж этой книги, она появится в главах про страхи и про мошенников. Её первая мысль о нейросети была не «интересно, как это работает», а «сколько это будет мне стоить и не выманят ли у меня данные». Мысль здравая. По данным ВЦИОМ, среди россиян старше шестидесяти нейросетями пользуются тридцать девять процентов, и еженедельно из них — меньше четверти. Не потому, что не умеют нажимать кнопки. Потому что не понимают, во что ввязываются, и никто не объяснил.

Эта книга написана так, чтобы Татьяна дочитала.

О чём эта книга и чего в ней не будет

Я пользуюсь этими инструментами с 2023 года: в работе с текстами и в книгах. Последние полтора года — ещё и в музыке, которую делаю с нейросетями и выкладываю под тем же именем, что стоит на обложке. Опыт разный, и в основном он состоит из неудач, разобранных по косточкам. Про удачные случаи писать нечего: они выглядят как «попросил и получил».

Поэтому книга устроена так.

Сначала — как это работает на самом деле, без единого термина из машинного обучения. Пять глав, после которых вы будете понимать, почему нейросеть врёт, что она знает и чего не знает.

Потом — как начать сегодня: какие сервисы работают в России, сколько это стоит, что писать в пустое окно, какие приёмы формулировки подтверждены измерениями, а какие оказались фольклором.

Третья часть — самая для меня важная: как проверять. Ни одна книга по этой теме на русском языке, из тех четырнадцати, что я разобрал перед тем как сесть за эту, не отвечает на вопрос «а вдруг она соврала». Отвечать буду я.

Четвёртая — про жизнь: дом, работа, деньги, здоровье, дети, мошенники. Пятая — про то, чего бояться и что из этого подтверждается данными.

Чего в книге не будет. Устройства моделей изнутри: вам не нужно знать, что такое трансформер, чтобы понять, почему модель уверенно врёт. Программирования. Способов обойти ограничения и оплатить зарубежные подписки — это отдельный юридический сюжет, и я его касаться не буду. Обещаний, что ИИ заменит вам юриста, врача и бухгалтера: там, где такие обещания сбывались, люди попадали в больницу, и об этом тоже будет глава.

И ещё одного не будет — восторга. Ниша им перегрета, вы на него уже один раз купились.

Четыре вечера, которые я предлагаю потратить

Прежде чем идти дальше, договоримся о цене вопроса. Книга просит у вас не курса и не месяца обучения — она просит четырёх вечеров по полчаса, разбросанных по ближайшим неделям.

Первый вечер уходит на то, чтобы завести один сервис и понять, где у него что: где включается поиск, где прикрепляется файл, где начинается новый разговор. Полчаса, и больше к этому возвращаться не придётся.

Второй — на одну настоящую задачу, доведённую до результата. Не «попробовать», а сделать дело, которое всё равно надо было сделать.

Третий — на проверку: научиться за две минуты убеждаться, что вам не соврали. Это единственный навык, без которого пользоваться этим инструментом по-настоящему опасно, и ему посвящена целая часть книги.

Четвёртый — на разговор с близкими. Про то, что нельзя загружать, и про кодовое слово на случай звонка знакомым голосом. Звучит странно в первой главе; к разговору про поддельные голоса вы поймёте, почему я ставлю это в один ряд с остальным.

Всё. Двух часов достаточно, чтобы инструмент перестал быть игрушкой из новостей и стал чем-то вроде калькулятора: вы не думаете о нём, но пользуетесь ежедневно.

Я обещаю не превращать эти два часа в мучение. В книге не будет домашних заданий на вечер, схем с квадратами и призывов выйти из зоны комфорта. Будет то, что работает, с указанием, откуда я это знаю, — и прямое признание там, где не знаю.

И ещё одно, последнее в этой главе. Если после всех попыток вы решите, что вам это не нужно, — такой исход тоже нормальный. Есть работа и есть жизнь, где выигрыш от этого инструмента невелик, и дальше я покажу, как это измерить секундомером, а не почувствовать. Отказ, основанный на замере, — взрослое решение. Отказ, основанный на одном неудачном обеденном перерыве, — просто неудачный обеденный перерыв.

Что унести из главы

Ваш первый неудачный опыт — это статистика, а не приговор. Доля разочарованных растёт вместе с числом пользователей: почти четверть россиян считает инструмент бесполезным для своих задач, и за год эта доля выросла. Говорит она не о людях, а о том, как инструмент показывают.

Обрывается всё в трёх точках: недоступный сервис, пустое окно, водянистый первый ответ. Все три чинятся, и ни одна не требует технических знаний.

Проверьте на себе прямо сегодня. Вспомните последнюю задачу, которую решали руками дольше получаса, — письмо, таблицу, разбор документа, поиск информации. Не задавайте её нейросети. Пока просто запишите на бумаге в одну строку: что было, чего вы хотели добиться,

кто должен был это прочитать. Через семь глав мы к этой записи вернёмся, и вы увидите разницу между тем, что вы спросили в первый раз, и тем, что стоило спросить.

А начнём мы с вопроса, который обычно пропускают: что вообще происходит внутри, когда вы нажимаете «отправить».

Глава 2. Что такое нейросеть на самом деле

Откуда она берёт ответ?

Вопрос кажется наивным, но именно от ответа на него зависит всё дальнейшее: почему она врёт, почему соглашается, почему не знает вчерашних новостей и почему на один и тот же вопрос отвечает по-разному. Обычно на этот вопрос отвечают либо словами «искусственный интеллект анализирует огромные массивы данных», что не значит ничего, либо схемой с нейронами и слоями, после которой человек закрывает книгу.

Я попробую третий способ. Без единого термина из машинного обучения и без вранья.

Она не ищет ответ. Она его дописывает

Начнём с того, чего не происходит. Внутри нейросети нет базы знаний, куда она заглядывает. Нет справочника, нет таблицы фактов, нет поиска по интернету — если поиск не подключён отдельной функцией. Когда вы отправляете вопрос, она не находит готовый ответ. Она его составляет.

Составляет по одному принципу: какое слово вероятнее всего идёт следующим.

Яндекс Практикум в июле 2026 года сформулировал это точнее всех, что я встречал на русском: модели не проверяют факты так, как это делает человек, а прогнозируют наиболее вероятное продолжение текста. Всё. Это полное описание механизма.

Вы пишете «столица Франции — », и модель продолжает «Париж», потому что в текстах, на которых её учили, после этих слов почти всегда шёл Париж. Вы пишете «напиши письмо в управляющую компанию», и она пишет самое вероятное письмо в управляющую компанию: с обращением «уважаемые», с просьбой рассмотреть обращение, без вашей текущей протечки.

Она не знает, что Париж — столица. Она знает, что здесь идёт слово «Париж». Разница выглядит философской ровно до того момента, когда модель с той же уверенностью допишет несуществующий номер закона.

T9 на стероидах и почему эта метафора хромает

Есть метафора, которая для человека тридцати пяти — шестидесяти лет работает лучше любой другой. В марте 2023 года Павел Комаровский и Игорь Котенков выпустили разбор «Эволюция нейросетей от T9 до ChatGPT» — до сих пор, на мой взгляд, лучший русскоязычный текст на эту тему. Их формулировка: это T9 из вашего телефона, только на стероидах.

T9 вы помните. Кнопочный телефон, вы набираете начало слова, он предлагает продолжение. Иногда угадывает, иногда подставляет глупость, и вы отправляете сообщение с чужим словом. Никакого понимания там не было — была статистика: какие слова чаще встречаются вместе.

Языковая модель делает то же самое, только предсказывает не следующее слово в слове, а следующий кусочек в тексте любой длины: в письме, в статье, в ответе на ваш вопрос. И статистику она собрала не с вашего телефона, а с такого объёма текста, который человек не прочитает за десять жизней.

Метафора хорошая. Но у неё есть слабое место, и я обязан его назвать сразу, иначе она вас обманет.

T9 предсказывал плохо, и вы это видели. Модель предсказывает так хорошо, что результат перестаёт быть похож на угадывание. Она пишет связный текст, держит стиль, ловит иронию, разбирает вашу таблицу, объясняет теорему. Из «это просто угадывание следующего слова» очень легко сделать вывод «значит, это ерунда», и вывод будет неверным.

Правильный вывод другой: это угадывание, которое доведено до такого качества, что часто неотличимо от знания. Часто — но не всегда. И вся практическая часть этой книги про то, как отличать.

Насмотренность вместо правил

Второе слово, которое стоит забрать в свой словарь, — насмотренность. Оно из речи дизайнеров и фотографов: человек посмотрел десять тысяч хороших обложек и теперь чувствует, что здесь шрифт не тот, хотя правила назвать не может.

Модель учится так же. Её не программировали правилами русского языка. Ей не объясняли, что такое падеж и почему после «согласно» идёт дательный. Она увидела столько текста, что научилась воспроизводить его закономерности сама. Яндекс Практикум описывает это тем же словом: нейросеть учится путём насмотренности, собирая массивы информации и выявляя закономерности.

Отсюда две вещи, которые дальше объяснят половину странностей.

Первая: у модели нет источника. Она не помнит, откуда узнала, что аспирин разжижает кровь. Знание размазано по её устройству так же, как у вас размазано умение говорить по-русски: вы не помните, из какой книги взяли слово «навзничь».

Вторая: если чего-то в текстах было мало, модель это воспроизведёт плохо — но всё равно уверенно. Редкая фамилия, местный закон, тарифы вашей управляющей компании, свежая новость. Она не скажет «мало данных». Она допишет самое вероятное. Правдоподобное, а не правдивое.

Есть ещё третья метафора, расхожая: нейросеть похожа на мозг, она учится на своих ошибках. Словарь терминов, который в марте 2026 года опубликовал MTS Media, описывает её примерно так — миллионы виртуальных нейронов передают друг другу сигналы. Технически придаться не к чему, но для читателя это опасная метафора: она незаметно подсказывает, что машина понимает. Понимания там нет. Есть очень хорошая статистика.

Три слова и две даты, которые снимают половину путаницы

Теперь короткий, но важный кусок про слова. Без него вы будете спотыкаться в каждой второй новости.

Нейросеть — это способ устройства программы. Не только языковой: нейросети распознают лица, ставят диагнозы по снимкам, предсказывают погоду. В обиходе слово захватило всё: «нейронка», «нейросеть написала», «спрошу у нейросети». В опросах ВЦИОМ так и пишут — «нейросети в нашей жизни». Это слово вашего собеседника, и я буду им пользоваться.

Модель — конкретный экземпляр, обученный и выпущенный: GPT-5, Gemini 3, GigaChat, YandexGPT. У модели есть версия и дата выпуска, как у прошивки. Модель — то, что думает.

Сервис — то, куда вы заходите: приложение, сайт, окно в мессенджере. ChatGPT, Алиса AI, ГигаЧат. Сервис — то, где вы разговариваете.

Путаница между вторым и третьим стоит людям нервов. Человек читает новость «вышла новая модель», идёт в своё приложение и не видит разницы — потому что сервис ещё не переключили. Или наоборот: сервис обновили, а человек уверен, что говорит с той же программой, что месяц назад, и не понимает, почему ответы изменились.

Практический вывод простой. Когда вам говорят «ИИ теперь умеет генерировать видео» — спрашивайте, какой именно сервис и в каком тарифе. Умение принадлежит не «искусственному интеллекту вообще», а конкретному продукту в конкретном месяце.

Она думает не словами, а кусочками

Ещё одно слово, без которого не обойтись, потому что вы встретите его в тарифах: токен.

Модель работает не со словами и не с буквами, а с кусочками текста. Короткое частое слово — один кусочек. Длинное или редкое разваливается на два-три. Для русского языка гру-

бая прикидка такая: один токен — примерно три четверти слова, а страница обычного текста — это порядка семисот-восьмисот токенов.

Зачем вам это знать? Затем, что в токенах измеряется всё, за что берут деньги, и всё, что ограничивают. Сбер, например, даёт физическим лицам бесплатный годовой объём в токенах — на сентябрь 2026 это триста шестьдесят пять миллионов на все модели GigaChat вместе. Цифра выглядит астрономической, и по бытовым меркам она такая и есть: обычный человек за год не выберет и малой доли.

В токенах же измеряется и объём разговора, который модель удерживает. У неё нет памяти в человеческом смысле — есть окно, в которое влезает переписка. Всё, что не влезло, для неё перестаёт существовать. Отсюда знакомая многим история: на двадцатом сообщении модель забывает, о чём вы договаривались на третьем. Это не сбой и не «она устала». Это край окна.

Практический вывод один и простой: длинный разговор о разном — плохая идея. Новая задача — новый чат.

Почему она заговорила по-русски именно сейчас

Короткая справка, чтобы вы понимали, почему всё случилось так быстро.

Публичный запуск ChatGPT — тридцатое ноября 2022 года. До этого языковые модели существовали лет пять, но жили внутри исследовательских лабораторий и разговаривать с ними мог не всякий программист. Изменилось не устройство, а дверь: модель посадили в окно чата, и оказалось, что этого достаточно.

Русскоязычные сервисы появились через полгода. Двадцать четвёртого апреля 2023 года Сбер представил GigaChat. Семнадцатого мая того же года Яндекс построил YandexGPT в Алису — и, к чести компании, сразу предупредил, что Алиса может ошибаться в фактах. Предупреждение это тогда почти никто не заметил, а зря: в нём вся суть дела.

С тех пор поколения моделей сменялись примерно раз в год, и каждое следующее делало то, что предыдущее не умело: разбирать картинки, слушать голос, работать с файлами, искать в интернете. Я вернусь к этой хронологии в последней главе, когда буду объяснять, как не отстать. Пока достаточно одного: механизм всё это время оставался тем же самым. Модель по-прежнему достраивает текст по вероятностям. Она просто делает это точнее.

Почему один и тот же вопрос даёт разные ответы

Это первое, что сбивает с толку новичка, и объясняется тем же механизмом.

Модель предсказывает не одно слово, а вероятности для многих слов сразу. Дальше она выбирает — и выбирает не всегда самое вероятное, иначе тексты выходили бы деревянными и одинаковыми. Есть настройка случайности; в сервисах для обычных людей её обычно не показывают, но она работает всегда.

Отсюда следствие, которое пригодится вам при проверке. Если вы зададите один и тот же фактический вопрос трижды в разных чатах и получите три разных ответа — это сигнал. Не доказательство ошибки, но сигнал: модель здесь не уверена, она достраивает. Если три раза придёт одно и то же — уверенность выше. Не гарантия, но выше. Этим приёмом, кстати, пользуются и сами разработчики: Anthropic в своей документации по снижению выдумок прямо советует прогнать запрос несколько раз и сравнить ответы.

Второе следствие бытовое: не пугайтесь, если ответ у вас и у коллеги не совпал дословно. Это не значит, что кому-то из вас досталась «плохая версия».

Что из этого следует практически

Соберу механику в четыре следствия, которые пригодятся вам уже завтра.

Следствие первое: она не проверяет, она вспоминает похожее. Значит, вопрос «а это точно так?» бессмыслен по устройству — модель не может сходить и посмотреть, если ей не дали поиск. Полезнее спрашивать не о правильности, а об основании: откуда это, на чём держится, есть ли источник.

Следствие второе: чем чаще тема встречалась в текстах, тем надёжнее ответ. Из этого выводится карта доверия, которой я пользуюсь ежедневно. Общеизвестное — надёжно. Профессиональное, но распространённое — обычно надёжно. Редкое, местное, свежее, ваше личное — ненадёжно всегда. Именно поэтому она прекрасно объяснит, как работает страхование ответственности, и сойдёт про тариф вашей страховой компании.

Следствие третье: форма даётся ей легче содержания. Модель безупречно воспроизводит структуру делового письма, вид ссылки, стиль искового заявления, интонацию отзыва. Наполнение при этом может быть вымышленным. Отсюда обманчивое впечатление компетентности: мы оцениваем текст по форме, а форма у неё всегда отличная.

Следствие четвёртое: качество ответа зависит от количества полезного, что вы дали. Не от вежливости, не от волшебных слов, не от тарифа. От того, сколько определённости вы положили в вопрос. Об этом вся вторая часть книги, и это самый управляемый параметр из всех.

Есть и пятое следствие, о котором говорят реже. Раз модель воспроизводит закономерности текстов, она воспроизводит и их перекосы. Если в текстах какая-то точка зрения встречалась чаще, она будет чаще звучать и в ответах — не потому, что она вернее, а потому, что её было больше. На бытовых вопросах это незаметно, на спорных — вполне. Хороший способ это увидеть: задать один и тот же спорный вопрос с противоположных позиций и сравнить два ответа.

Ирина из первой главы, когда я объяснил ей механизм на примере Т9, сказала фразу, которую я потом повторял многим: «То есть она мне не отвечает, а сочиняет самое подходящее». Да. И качество сочинения зависит от того, сколько подходящего вы ей дали.

Попробуйте сегодня одно упражнение на пять минут. Задайте любому доступному вам чату вопрос, ответ на который вы точно знаете: про свой город, свою профессию, свой дом. Потом задайте тот же вопрос ещё раз в новом чате. Сравните два ответа между собой и с тем, что знаете вы. Вы за пять минут увидите и силу инструмента, и его слабое место — раньше, чем я успею про них рассказать.

А в следующей главе разберём, почему это слабое место не чинится и почему модель врёт не по ошибке, а по устройству.

Глава 3. Почему она врёт и почему это не поломка

Семьдесят пять процентов.

Это доля неверных утверждений, которые выдала одна из моделей OpenAI на наборе фактических вопросов под названием SimpleQA. Цифру опубликовала сама OpenAI в сентябре 2025 года в работе с прямым заголовком: «Почему языковые модели галлюцинируют». Компания, которая продаёт эти модели, честно написала, что её модель на этом наборе вопросов ошибается в трёх ответах из четырёх.

И в той же таблице стоит вторая цифра, ради которой всё и затевалось. Другая модель того же семейства ошибается в двадцати шести процентах случаев. Не потому, что она в три раза умнее. Потому что она в половине случаев отказывается отвечать.

Разберём эту пару чисел медленно. В ней объяснение всего, за что нейросети ругают.

Модель, которая молчит, проигрывает модели, которая врёт

Вот полная картина из той работы. Модель o4-mini: воздерживается от ответа в одном проценте случаев, отвечает правильно в двадцати четырёх, ошибается в семидесяти пяти. Модель GPT-5-thinking-mini: воздерживается в пятидесяти двух процентах, отвечает правильно в двадцати двух, ошибается в двадцати шести.

Правильных ответов почти поровну — двадцать четыре против двадцати двух. А неверных втрое меньше у той, которая умеет промолчать.

Теперь главное. Почему первая не молчит?

Объяснение, которое даёт OpenAI, я считаю самым полезным фактом во всей этой книге. Модели оценивают по доле правильных ответов. Не по доле честных. Если модель на экзамене из ста вопросов знает шестьдесят и на сорок отвечает «не знаю», она получает шестьдесят баллов. Если она знает те же шестьдесят, а на сорок пишет правдоподобное угадывание, часть угадываний попадёт — и балл будет выше.

Вы делали то же самое в школе. На тесте с четырьмя вариантами вы не оставляли пустое поле, а ставили галочку наугад, потому что за пустое поле точно ноль, а за галочку есть шанс. Модель обучали ровно этой логике, тысячами прогонов, всё время её создания.

Отсюда вывод, который переворачивает отношение к инструменту: выдумка — не сбой модели, а её натренированное поведение. Она не сломалась, когда сочинила несуществующий закон. Она сделала то, за что её всю жизнь хвалили: не оставила поле пустым.

Слово «галлюцинация» вам мешает

Термин прижился, и я буду им пользоваться, потому что вы встретите его везде. Но он неудачный, и вот чем.

Галлюцинация у человека — состояние необычное, болезненное, отличное от нормы. Слово подсказывает, что модель обычно работает правильно, а иногда сбоит. На самом деле механизм один и тот же всегда: она достраивает вероятное продолжение. Когда данных было много, вероятное продолжение совпадает с правдой. Когда мало — не совпадает. Изнутри процесса разницы нет никакой.

Точнее звучит формулировка, которую я слышал в обсуждениях на профессиональных площадках: она уверенно врёт. Не «иногда сбоит», а именно так — с ровной интонацией, без единого признака сомнения. В комментариях к разбору проверок моделей в январе 2026 года это сформулировали жёстче: машина ошибается с полной уверенностью в голосе, и это опаснее человеческого вранья. Человек, который не знает, обычно мнётся. Модель — нет.

Где она врёт чаще всего

За три года у меня накопился список мест, где выдумка почти гарантирована. Это не исследование, а записи практика, и я так их и подаю.

Ссылки и адреса страниц. Самое частое. Модель прекрасно знает, как выглядит ссылка, и составляет правдоподобную: правильный домен, правдоподобный раздел, осмысленный набор слов в конце. Страницы по ней не существуют. Проверяется за две секунды и потому обиднее всего.

Цитаты. Особенно цитаты известных людей. Модель знает стиль, знает темы, знает, что этот человек мог бы такое сказать, — и составляет фразу, которой никто не говорил.

Номера и даты в документах. Номер закона, пункт договора, номер судебного дела, дата вступления в силу. Форма правильная, содержание выдуманно. Это место, где люди теряют деньги: российский суд уже штрафовал за жалобу со ссылками на несуществующие судебные акты.

Фамилии и должности. Кто автор исследования, кто возглавляет ведомство, кто написал книгу. Модель соединяет правдоподобных людей с правдоподобными делами.

Ваша местная конкретика. Тарифы вашей управляющей компании, расписание вашей поликлиники, условия вашего банка. Этого в её текстах не было никогда, но признаться она не умеет.

Общий признак у всех пяти пунктов один: это факты, у которых есть строгая форма и мало упоминаний. Форму модель воспроизводит блестяще. Содержание берёт из воздуха.

И зеркальное наблюдение, полезное практически: там, где тема обсуждалась миллион раз, выдумок почти нет. Как варить гречку, чем отличается ОСАГО от каско, что делать при ожоге первой степени — здесь модель надёжна, как приличный справочник. Опасность не размазана ровным слоем. Она сидит в редком.

Лишняя фраза, которая ломает решение

Есть исследование, которое я люблю за наглядность. В октябре 2024 года исследовательское подразделение Apple опубликовало работу с набором задач GSM-Symbolic. Брали простые школьные задачи по математике и добавляли в условие одну фразу — осмысленную, но никак не влияющую на решение. Что-то вроде: «пять киви были мельче обычных».

Точность моделей падала — иногда до шестидесяти пяти процентов от исходной.

Человек эту фразу пропускает: он понимает, что размер киви не влияет на их количество. Модель не понимает, она сопоставляет с образцом. Появилось лишнее число или лишнее уточнение — образец поплыл.

Здесь я обязан сделать оговорку, которую в популярных пересказах не делают. В 2026 году этот же набор задач прогнали на новых моделях, и результаты вышли заметно лучше. Значит, точная формулировка не «нейросети не умеют думать», а другая: устойчивость их рассуждения ниже, чем кажется по гладкости ответа, и она растёт от поколения к поколению. Тот, кто цитирует работу Apple как доказательство глупости моделей, цитирует позапрошлый год.

Практический вывод для вас прямой: чем чище условие, тем надёжнее ответ. Лишние подробности в запросе не украшают его, а сбивают.

Что меряют эти проценты и чего они не меряют

Теперь заступлюсь за инструмент, иначе картина выйдет однобокой.

Цифра «ошибается в семидесяти пяти процентах» звучит как приговор, но она не про вашу жизнь. SimpleQA — специально составленный набор коротких вопросов, подобранных так, чтобы модели на них спотыкались. Это тест на прочность, а не замер бытовой точности. Если бы модель врала в трёх ответах из четырёх на обычных вопросах, ею не пользовались бы семьдесят восемь процентов российских интернет-пользователей.

Так устроены все подобные измерения. Они отвечают на вопрос «где инструмент ломается», а не «как он работает в среднем». Когда вы в новостях увидите очередной процент, первым делом спрашивайте: на каких вопросах меряли.

Отдельно скажу про то, чего эти проценты не измеряют вовсе, — про интонацию. Ни один известный мне бенчмарк не оценивает, насколько уверенно звучит неверный ответ. А это как раз то, что вредит больше всего: в измеренных экспериментах ответы, сделанные с ИИ за пределами его возможностей, оказывались убедительнее правильных.

Есть, кстати, и хорошая новость в том же исследовании OpenAI. Раз модель молчит чаще, когда ей это разрешают, значит, разрешение работает. Прямая просьба «если не уверен — скажи, что не знаешь» действительно повышает долю честных отказов. Anthropic в своей документации ставит этот приём на первое место среди способов уменьшить выдумки. Это не волшебное заклинание и не гарантия, но одна фраза, которая заметно меняет поведение, стоит того, чтобы её запомнить.

Почему подключённый поиск помогает не до конца

Резонное возражение: если модель не знает свежего и врёт в редком, пусть посмотрит в интернете. Многие сервисы это умеют — Алиса, ГигаЧат, поисковые ответы Яндекса и Google.

Помогает, и заметно. В обсуждениях за 2026 год встречается ровно эта формулировка: с включённым поиском ответы надёжнее, без него — выдумки. Логика понятна: модель не достраивает по памяти, а берёт кусок реального текста и пересказывает.

Но проблема не исчезает, а меняет вид. Появляются две новые.

Первая: поиск нашёл не то. Модель добросовестно пересказала страницу, которая сама содержит ошибку или устарела на два года. У неё нет способа оценить, хороший ли это источник, — она видит текст, а не репутацию.

Вторая, более коварная: ссылка настоящая, а пересказ вольный. Модель ставит рядом с абзацем адрес реальной страницы, и человек расслабляется — ссылка же есть, значит, проверено. А в абзаце сказано не совсем то, что на странице. Или то, но с чужой цифрой.

Отсюда правило, которое я вывел для себя и которым пользуюсь до сих пор: наличие ссылки в ответе не означает, что ответ ей соответствует. Ссылка — это приглашение проверить, а не отчёт о проверке. Открывать её занимает секунды, и дальше в книге будет короткая процедура, которая укладывается в две минуты.

Что с этим делать: короткий ответ

Полный ответ — вся третья часть книги, шесть глав. Здесь дам рабочий минимум, чтобы вы могли пользоваться инструментом уже завтра.

Первое. Разделите свои вопросы на два вида. Вопросы на понимание — объясни, перекажи, помоги сформулировать, придумай варианты, разбери мой текст. Здесь выдумка почти не страшна: вы сами видите, годится ответ или нет. И вопросы на факт — сколько, когда, кто, какой номер, где написано. Вот здесь начинается работа.

Второе. Всё, что вы собираетесь передать дальше — начальнику, клиенту, врачу, в документ, — проверяется независимо от того, насколько уверенно оно звучит. Не потому что модель плохая, а потому что отвечать за это будете вы. Одному представителю компании пропущенная проверка стоила пятидесяти тысяч рублей штрафа.

Третье. Разрешайте ей не знать. Одна строка в конце запроса: «если данных нет, так и напиши, не придумывай». Это не гарантия, но измеренная разница между моделью, которая молчит, и моделью, которая всегда отвечает, — в три раза по числу ошибок.

Ирина, когда я объяснил ей всё это, спросила ровно то, что спрашивают все: «Тогда зачем оно нужно, если врёт?» Ответ такой же, как про любой инструмент с известной погрешностью. Микроволновка не готовит стейк. Это не делает её бесполезной — это делает её микроволновкой. Проблема начинается там, где человек ждёт от неё стейка и обижается.

Как выглядит выдумка вблизи

Полезно один раз рассмотреть её в упор, чтобы дальше узнавать без объяснений.

Возьмём типичный случай: вы спрашиваете про порядок оформления какой-нибудь бытовой процедуры — возврата товара, оформления пособия, подачи показаний. Ответ приходит стройный: сначала то, потом это, срок такой-то, основание — приказ такой-то от такого-то числа.

Первые две трети ответа, скорее всего, верны. Процедура описана правильно, потому что про неё написаны тысячи статей. А вот номер приказа и дата — самое слабое место, потому что в текстах они встречались редко и в разных вариантах.

Заметьте, что выдумка не выглядит как выдумка. Она не написана другим шрифтом и не сопровождается оговоркой. Она стоит в том же ровном ряду, что и правда, и берёт достоверность взаймы у соседей по предложению.

Это и есть главная механика обмана — не злого умысла, а именно обмана восприятия. Верное соседство делает неверное правдоподобным. Человек читает: «сначала подаёте заявление, потом ждёте десять дней, основание — приказ номер такой-то». Первые две части он может проверить своим опытом и убеждается, что всё так. Третью проверить не может — и по инерции доверяет.

Отсюда практический навык, который стоит выработать: читать ответ не целиком, а по кускам, отделяя проверяемое от объяснительного. Объяснительная часть — как устроено, в каком порядке, что за чем — обычно надёжна. Точечная — номер, дата, фамилия, сумма, ссылка — ненадёжна всегда.

Второй признак, который замечают редко: чем конкретнее вопрос про редкое, тем правдоподобнее выдумка. Спросите про закон вообще — получите общие слова. Спросите номер и пункт — получите номер и пункт. Точность вопроса не добавляет модели знаний, она добавляет уверенности в ответе.

И третье, что стоит знать: одна и та же модель в двух разных чатах на редкий вопрос часто отвечает по-разному, а на распространённый — одинаково. Это готовый бесплатный тест, и он ничего не стоит, кроме минуты.

Что унести из главы

Модель врёт не из-за поломки. Её обучали и оценивали так, что молчание проигрывает угадыванию: воздерживаться от ответа — значит терять баллы. Отсюда разница между моделью с одним процентом отказов и семьюдесятью пятью процентами ошибок и моделью с пятьюдесятью двумя процентами отказов и двадцатью шестью процентами ошибок.

Выдумка живёт в редком и в строгом по форме: ссылки, цитаты, номера, фамилии, ваша местная конкретика. В обкатанных бытовых темах модель надёжна.

И самое неприятное: неверный ответ звучит ровно так же, как верный. Никакого сигнала не будет.

Проверьте сегодня сами, это займёт три минуты. Спросите у любого доступного вам чата ссылку на конкретный документ — например, на страницу вашей поликлиники с расписанием врачей. Получите ответ и откройте ссылку. Дальше вы всё поймёте про эту главу лучше, чем из моего текста.

Глава 4. Двенадцать слов, которые нужно знать

— У тебя контекст переполнился, начни новый чат.

— Что переполнилось?

— Ну, окно. Она уже не помнит начало.

— Какое окно, я в приложении сижу.

Этот разговор Ирина пересказала мне почти дословно: так ей объясняла дочь. Дочь не виновата, она говорила на нормальном языке — просто на языке, которого мать не знает. А мать после такого разговора чувствует себя глупее, чем была до него, и это худшее, что может случиться с человеком, который только начал.

Словарь темы выглядит устрашающе. В словаре терминов искусственного интеллекта, который MTS Media опубликовал в марте 2026 года, несколько десятков статей. Большинство из них вам не понадобится никогда.

Рабочий минимум — двенадцать слов. Я проверял: этого хватает, чтобы не спотыкаться ни в одном интерфейсе и понимать девять из десяти новостей по теме.

Четыре слова про то, с чем вы разговариваете

Искусственный интеллект, ИИ. Слово-зонтик. Им называют всё: и программу, которая подбирает вам музыку, и камеру на перекрёстке, и чат, куда вы пишете вопросы. Само по себе оно почти ничего не сообщает. Когда в новости написано «ИИ научился», первым делом выясняйте, что именно за программа и кто её сделал.

Нейросеть. Способ устройства программы: она не следует записанным правилам, а обучается на примерах. Слово в обиходе захватило всё подряд — «нейронка», «спрошу у нейросети». Так говорят люди, так пишут в опросах ВЦИОМ, так буду писать и я.

Модель. Конкретная обученная программа с именем и версией: GPT-5, Gemini 3, GigaChat, YandexGPT. Модель — то, что отвечает. У неё есть дата выпуска, и она важнее, чем кажется: модель прошлого года и модель этого — разные собеседники.

Чат-бот, сервис. Окно, в которое вы пишете. ChatGPT, Алиса AI, ГигаЧат — это сервисы; внутри у них модели, которые владелец сервиса может поменять, не спросив вас. Отсюда бытовое наблюдение: если ответы вдруг стали другими, вы, скорее всего, не сошли с ума, а попали на обновление.

Разница между моделью и сервисом кажется занудством ровно до первого спора. Человек говорит «ChatGPT не умеет считать», ему возражают «умеет, я вчера считал», и оба правы: у одного бесплатный тариф с быстрой моделью, у другого платный с медленной и думающей.

Четыре слова про то, как она отвечает

Промпт, он же запрос. То, что вы пишете в поле ввода. Слово прижилось в русском без вариантов: оно стоит в словарях терминов без кавычек и без перевода. Я буду говорить и «промпт», и «запрос» как синонимы. Важнее слова смысл: промпт — это не вопрос, а техническое задание. Разница между вопросом и заданием — тема восьмой и девятой глав.

Токен. Кусочек текста, которым модель меряет всё: и то, что вы написали, и то, что она ответила. Примерно три четверти слова для русского языка; страница текста — около семисот-восемисот токенов. Слово вам понадобится только в одном месте — в тарифах. Когда Сбер пишет, что физическому лицу бесплатно доступно триста шестьдесят пять миллионов токенов в год на все модели GigaChat, речь именно об этих кусочках.

Контекстное окно. Объём переписки, который модель держит перед глазами, пока отвечает. Всё, что не влезло, для неё перестаёт существовать — не «забылось», а именно перестало существовать. Отсюда та самая фраза дочери про переполнившийся контекст: на длинном разговоре модель теряет начало.

Полезная бытовая аналогия: это не память, а стол. Всё, что лежит на столе, модель видит. Что не поместилось — упало на пол, и поднимать она не умеет. Отсюда практическое правило, к которому я ещё вернусь: длинный разговор портится, и новая задача заслуживает нового чата.

Галлюцинация. Уверенно выданная выдумка: несуществующая ссылка, придуманная цитата, вымышленный номер закона. Термин неточный, я разобрал это в предыдущей главе, но он общепринят. Если запомнить из всего словаря одно слово, пусть будет это.

Четыре слова про то, что она умеет

Мультимодальность. Способность работать не только с текстом: разбирать картинку, слушать голос, читать файл. Слово громоздкое, встречается в описаниях сервисов. На практике оно означает, что вам не обязательно печатать — можно сфотографировать или продиктовать. Двенадцатая глава про это целиком.

Генерация. Создание нового: текста, картинки, музыки, видео. «Сгенерить» — глагол, который вы услышите от всех, кто пользуется. Важная деталь, о которой узнают поздно: у сгенерированного другой правовой статус, чем у снятого или нарисованного. В условиях Шедеврума, например, в редакции апреля 2026 года прямо сказано, что созданные изображения пользователь вправе применять в личных некоммерческих целях, а коммерческое использование требует согласования. Мелкий шрифт, который стоит прочесть до того, как вы поставите картинку на обложку.

Агент. Модель, которая не только отвечает, но и действует: сама ходит по сайтам, заполняет формы, собирает данные в несколько шагов. Первые массовые агентские режимы появились в 2025 году, у Яндекса такие функции встроены в Алису. Для вас это слово означает одно: если программа делает что-то за вас без вашего подтверждения на каждом шаге, ошибки она тоже сделает без подтверждения.

Лимит. Сколько вам можно бесплатно. Ни OpenAI, ни Яндекс, ни Сбер не публикуют постоянных чисел вроде «двадцать сообщений в день»: лимиты меняются без объявления и зависят от нагрузки. Все конкретные цифры, которые вы встретите в интернете, — чьи-то частные замеры. Относитесь к ним как к прогнозу погоды на послезавтра.

Пять слов второго ряда

Эти пять знать необязательно, но они попадают в интерфейсах, и лучше не пугаться.

Режим рассуждения. Настройка или отдельная модель, которая думает дольше и отвечает медленнее. Внутри она проговаривает решение по шагам, прежде чем выдать ответ. На арифметике, логике, разборе документов это заметно помогает; на «перепиши письмо покороче» — только тратит ваше время. В GPT-5, вышедшей в августе 2025 года, появился распределитель, который сам решает, когда думать дольше, а когда ответить сразу. Если у вас в интерфейсе есть переключатель — включайте его на задачах, где важна точность, и выключайте на болтовне.

Дата отсечки знаний. Число, после которого модель ничего о мире не знает. У GPT-5 при выходе седьмого августа 2025 года она была тридцатого сентября 2024 года — почти на год раньше. Следующая глава про это целиком, и это слово в ней главное.

Поиск с опорой на источники. Способ работы, при котором сервис сначала находит нужные страницы или ваши файлы, а потом просит модель ответить, глядя в них. Именно так устроены ответы поисковиков и разбор загруженного документа. В профессиональной речи это называют трёхбуквенным сокращением, вам оно не нужно.

Локальная модель. Модель, запущенная на вашем компьютере. Слабее облачной, требует приличного железа, зато работает без интернета и ничего никуда не отправляет. Для человека, которому важна тайна переписки, это единственный способ спать спокойно. Требования к железу в открытых источниках расходятся, и конкретных цифр я приводить не буду — их надо смотреть под конкретную модель в месяц покупки.

API. Способ подключить модель к другой программе, минуя окно чата. Вам это слово встретится в тарифах, где цены указаны за токены. Понимать его надо ровно в одном объёме: если вы читаете прайс-лист и не понимаете, почему там нет месячной подписки, — вы попали на страницу для разработчиков, идите обратно на страницу для людей.

Слова, которые вам не понадобятся

Теперь вторая половина работы — что можно смело пропускать мимо ушей. Список короче, чем кажется, но он снимает много тревоги.

Всё, что описывает устройство модели изнутри: трансформер, слои, веса, эмбединги, векторы. Это как знать, из чего сделан катализатор в вашей машине. Знание уважаемое, к вождению отношения не имеет.

Всё, что описывает создание моделей: обучение, дообучение, фэйн-тюнинг, датасет, квантизация. Работа инженеров.

Всё футурологическое: сильный ИИ, сверхинтеллект, сингулярность. Это разговор про будущее, в котором никто не разбирается, включая тех, кто в нём уверен.

И отдельно — промпт-инжиниринг как название профессии. В 2023 году так называли редкий навык, за который платили. В 2026-м это чаще всего вывеска курсов. Данные, из-за которых я так думаю, будут дальше, и они довольно обидные для продавцов курсов.

Есть ещё словечки, которые ничего не объясняют, но выдают своего. Вайб-кодинг, например: составление программ разговором с моделью вместо написания кода. Слово настолько разошлось, что словарь Collins в ноябре 2025 года назвал его словом года. Знать его вам незачем — но теперь вы понимаете, о чём речь, если услышите.

Слова, которые встретятся в новостях

Отдельно — четыре выражения, которые вам попадутся не в интерфейсе, а в заголовках. Знать их устройство не нужно, а понимать, о чём речь, полезно: в 2026 году тема ушла из технологического раздела в новостной.

Большие фундаментальные модели. Так называется предмет российского закона о поддержке развития технологий искусственного интеллекта, принятого в июле 2026 года. Термин задаёт порог — по сообщениям отраслевых изданий, речь о моделях не меньше миллиарда параметров. Для вас смысл один: закон адресован разработчикам и крупным площадкам, а не человеку, который пишет запрос в чат.

Суверенные и национальные модели. Категории из того же закона. Обе предполагают, что данные обрабатываются на территории России российскими юридическими лицами, и правительство вправе определять случаи, когда допустимо использовать только такие модели. В обычной жизни это, скорее всего, проявится в государственных сервисах и в корпоративных требованиях, а не в вашем телефоне.

Маркировка ИИ-контента. В законе она оформлена интересно: как право пользователя пометить созданный материал, и как обязанность площадок с суточной аудиторией больше пятисот тысяч человек дать пользователям техническую возможность такую пометку поставить. То есть маркировка становится нормой обихода, а не запретом.

Обучение на защищённых текстах. Тем же законом введено изъятие: использование правомерно полученных произведений для извлечения, сравнения и классификации при обучении моделей не считается нарушением авторских прав. Тема спорная во всём мире, и здесь у России теперь есть своя позиция.

Даты в этом абзаце я привожу с оговоркой: сроки вступления в силу отдельных норм в источниках расходятся, и перед тем как на них ссылаться, стоит свериться с официальным текстом. Это, кстати, хороший пример того, о чём вся книга: даже здесь я не могу быть уверен в детали, и лучше сказать об этом, чем написать красивую дату.

Как звучит человек, который правда пользуется

Небольшое наблюдение вместо резюме. Есть речевые приметы, по которым видно, что человек работает с этими инструментами, а не пересказывает статьи.

Он говорит «она придумала» вместо «произошла галлюцинация». Говорит «вылетел лимит», «ответ поплыл», «скормил ей файл», «переспросил другими словами». Жалуется, что модель «подтормаживает на длинном контексте». И почти никогда не называет себя экспертом по ИИ.

Обратная примета тоже надёжная. Если человек говорит «в эпоху нейросетей», «беспрецедентные возможности» и «революция искусственного интеллекта», он, скорее всего, продаёт курс. Проверять его слова стоит вдвое внимательнее.

Ирине из нашего разговора в начале главы не хватало ровно четырёх слов: модель, чат, контекстное окно, лимит. Через двадцать минут объяснений тот же диалог с дочерью прошёл спокойно. Дело было не в способностях. Дело было в словаре.

Что унести из главы

Двенадцати слов достаточно: искусственный интеллект, нейросеть, модель, сервис, промпт, токен, контекстное окно, галлюцинация, мультимодальность, генерация, агент, лимит. Остальные тридцать существуют для разработчиков и для продавцов курсов, а вам они нужны примерно так же, как устройство катализатора.

Из двенадцати самое полезное — контекстное окно и галлюцинация. Первое объясняет, почему длинный разговор портится. Второе — почему ответ нельзя принимать на веру.

Сделайте сегодня одну вещь: откройте свой сервис и найдите в нём глазами эти слова. Где-то будет написано про лимит, где-то про модель в настройках, где-то про загрузку файлов. Пять минут поиска по интерфейсу дают больше, чем страница объяснений: вы увидите, что все двенадцать слов уже лежат у вас перед носом, просто раньше были шумом.

Глава 5. Что она знает и чего не знает

Модель, с которой вы разговариваете сегодня, живёт в прошлом. Не в переносном смысле, а в буквальном: её знания заканчиваются на конкретной дате, и эта дата обычно на год старше, чем день, когда сервис появился у вас в телефоне.

Пример, который легко проверить. GPT-5 вышла седьмого августа 2025 года, а объявленная граница её обучающих данных — тридцатое сентября 2024 года. Одиннадцать месяцев разницы. Всё, что случилось в мире за эти одиннадцать месяцев, для модели просто не произошло: ни новых законов, ни новых цен, ни новых людей в должностях.

У этого свойства есть скучное название — дата отсечки знаний. И оно объясняет добрую половину случаев, когда человек говорит «она несёт чушь».

Почему граница вообще существует

Модель учат один раз. Это долгий и дорогой процесс: собирают гигантский объём текстов, прогоняют через него программу, получают на выходе набор настроек, который потом никак не меняется в разговоре с вами. Дальше модель проверяют, доводят, выпускают — и всё это занимает месяцы. Отсюда и разрыв между последним прочитанным текстом и днём выхода.

Дальше модель не читает. Она не подписана на новости, не следит за вашей отраслью, не узнаёт, что вчера поменялся тариф. Она законсервирована.

Отсюда первое практическое следствие. Любой вопрос вида «какая сейчас», «сколько сегодня стоит», «кто сейчас возглавляет» — вопрос за границей её знаний. Ответ придёт, и он будет уверенным, потому что модель не умеет говорить «этого я знать не могу». Она построит правдоподобное на основе того, что знала год назад.

Проверяется это за минуту. Спросите у своего сервиса, какое сегодня число, а потом — какое последнее крупное событие ему известно. Ответ вас удивит чаще, чем вы думаете.

Поиск — это отдельная пристройка

Резонное возражение: но она же отвечает мне про сегодняшние новости, я проверял.

Правильно. Потому что многие сервисы приделали к модели поиск. Это отдельная функция: сервис сначала ищет в интернете, находит несколько страниц, кладёт их текст модели прямо в контекст — на тот самый стол из прошлой главы — и просит ответить с опорой на них.

Такой поиск встроен в Яндекс: сервис «Нейро», который вместо списка ссылок даёт один ответ, запущен шестнадцатого апреля 2024 года. Google объявил об общедоступности ответов ИИ прямо в поисковой выдаче в мае 2024 года. У большинства чат-сервисов есть переключатель или он срабатывает автоматически.

Но заметьте, что произошло. Свежие данные знает не модель — их приносит поиск. Модель по-прежнему живёт в прошлом, просто ей на стол положили сегодняшнюю газету.

Для вас это означает две практические вещи.

Первая: узнайте, есть ли поиск в том сервисе, которым вы пользуетесь, и включён ли он. Одна и та же модель с поиском и без — два разных инструмента по надёжности на фактических вопросах.

Вторая: даже с поиском ответ не становится проверенным. Поиск может принести устаревшую страницу, а пересказ может разойтись с источником. Ссылка рядом с абзацем — приглашение проверить, а не отчёт о проверке.

Она не помнит вас — и не учится на вашем разговоре

Это, пожалуй, самое устойчивое заблуждение из всех, и оно живёт сразу в двух видах.

Первый вид — тревожный: «всё, что я напишу, немедленно уйдёт в её мозги, и завтра она расскажет это соседу». Второй — доверчивый: «я её обучил под себя, теперь она знает мои задачи».

Оба неверны, и по одной причине. Модель во время разговора не меняется. Её настройки зафиксированы при обучении, и ваш текст на них не влияет — ни один символ. То, что выглядит как «она запомнила», — это контекстное окно: пока переписка помещается на столе, модель видит всё сказанное и опирается на него. Закрыли чат — стол опустел.

Отдельно существует функция памяти между разговорами, она есть у части сервисов: система сохраняет заметки о вас и подкладывает их в начало новых чатов. Это тоже не обучение, а записка на столе. Её обычно можно посмотреть и стереть в настройках, и я рекомендую туда заглянуть — люди сильно удивляются тому, что там накопилось.

Теперь обоснованная половина тревоги. Ваши разговоры действительно могут использоваться — но не так и не тогда. По документации OpenAI, содержимое диалогов обычных пользователей может пойти на обучение будущих моделей, и это отключается в настройках приватности; для бизнес-тарифов и для подключения через программный интерфейс обучение по умолчанию не ведётся.

У российских сервисов свои формулировки, и они разные. В условиях использования Алисы AI в редакции от двадцать второго июня 2026 года прямо сказано, что Яндекс получает безвозмездную безотзывную лицензию на использование ваших запросов и загруженных материалов для улучшения алгоритмов и сервиса. В соглашении GigaChat для физических лиц в редакции от шестого февраля 2026 года, наоборот, оговорено, что контент клиента не используется в собственных целях, не связанных с предоставлением сервиса. Ни то ни другое соглашение не называет срок, через который загруженный файл удаляется.

Разница между «сейчас она учится на моих словах» и «через полгода мой текст может попасть в обучающий набор следующей модели» — принципиальная. Первое не происходит. Второе зависит от сервиса и от настроек, и это уже разговор про приватность — ему в книге отведена отдельная глава.

Чего она не знает никогда

Есть категория, которая не чинится ни поиском, ни новой версией модели. Это всё, чего не было в открытых текстах.

Ваши документы. Договор, который вы подписали. Приказ по вашей организации. Переписка с поставщиком. Если вы не загрузили файл в этот разговор, модель его не видит. И не скажет об этом — достроит типичный договор вместо вашего.

Вашу местную конкретику. Тарифы вашей управляющей компании, часы приёма вашей поликлиники, правила вашего садового товарищества, условия акции в вашем магазине. Это существует, но в текстах, до которых обучение не дотянулось.

Всё, что произошло после отсечки и не попало в поиск. Свежее решение суда, вчерашнее постановление, изменения в законе, вступившие в силу на прошлой неделе.

И вас. Ваш опыт, ваши обстоятельства, вашу цель, ваш срок, вашего начальника с его характером. Всё это придётся написать словами.

Я сформулировал для себя правило, которое оказалось на удивление рабочим: модель знает мир в целом и не знает ничего в частности. Она отлично объяснит, как устроен акт приёма-передачи, и не имеет понятия, что написано в вашем.

Проверка границы за одну минуту

Прежде чем доверять модели фактический вопрос, полезно понять, где у неё край. Это делается тремя короткими запросами, и я советую прогнать их в каждом сервисе, которым вы собираетесь пользоваться. Один раз.

Первый: «Какая у тебя дата отсечки знаний?» Хорошая модель называет месяц и год. Плохая уходит от ответа общими словами — это само по себе информация.

Второй: «Ты сейчас можешь искать в интернете? Если да, то как это включается?» Ответ покажет, поиск встроен постоянно, включается кнопкой или отсутствует. Заодно вы узнаете, как это выглядит именно в вашем интерфейсе, — а выглядит везде по-разному.

Третий: спросите про событие, которое случилось недавно и которое вы точно знаете. Не про мировое — про своё: изменение в вашей отрасли, местную новость, свежий закон, за которым следите. Дальше смотрите не на правильность, а на поведение. Модель призналась, что не знает? Хорошо. Уверенно рассказала неправду? Запомните это ощущение — оно вам ещё пригодится.

Занимает всё вместе минуту. Даёт понимание, которого не даст ни одна инструкция: у вашего инструмента появляется граница, и вы её видели своими глазами.

Кстати, о поколениях. Новая версия модели знает больше о мире, но ровно столько же о вас — то есть ничего. Обновление сдвигает дату отсечки и улучшает рассуждение; оно не добавляет модели знания о вашем договоре и вашей поликлинике. Люди часто ждут от новой версии не того, что она приносит, и потом разочаровываются на ровном месте.

Что с этим делать

Три рабочих привычки, которые снимают почти все проблемы этой главы.

Первая: разделяйте вопросы на общие и местные. Общие — как устроено, что это значит, каков порядок действий, чем отличается одно от другого — задавайте смело. Местные — сколько стоит у нас, что написано в моём договоре, когда работает вот эта контора — либо давайте вместе с документом, либо ищите глазами сами.

Вторая: приносите документ. Загруженный файл превращает местный вопрос в общий: теперь у модели на столе лежит ваш договор, и она разбирает его, а не типовой. Ограничения по форматам и размерам у российских сервисов вполне бытовые — Алиса AI принимает документы и PDF до ста мегабайт, GigaChat берёт текстовые файлы до сорока. Что при этом происходит с самим файлом — вопрос отдельный, и ответ там неприятный.

Третья: спрашивайте про дату. Простая строка в конце запроса — «на какую дату верна эта информация и откуда она у тебя» — экономит массу нервов. Модель либо честно скажет, что опирается на общие сведения, либо покажет источник из поиска. Оба ответа полезны, и оба лучше молчаливой уверенности.

Ирина проверила это на своём: спросила про изменения в порядке сдачи отчётности. Получила уверенный ответ с номером и датой. Открыла справочно-правовую систему, которой пользуется каждый день, и увидела, что порядок поменялся ещё весной. Модель об этом не знала — и не собиралась предупреждать.

Она сказала фразу, которую я потом цитировал многим: «То есть она как очень начитанный человек, который полгода просидел без интернета». Почти. С одной поправкой: этот человек ещё и не признаётся, что просидел.

Что сервис успел записать о вас

Раз уж речь зашла о памяти, стоит сделать одну вещь, которую почти никто не делает: посмотреть, что там накопилось.

У части сервисов есть функция, которая сохраняет заметки о пользователе между разговорами: чем вы занимаетесь, как вас зовут, что вы просили не делать, какие у вас предположения. Работает она незаметно — вы просто в какой-то момент обнаруживаете, что модель обращается к вам иначе, чем к незнакомцу.

Заметки эти обычно можно открыть в настройках, прочитать и стереть — целиком или по одной. Я советую заглянуть туда всем, кому этот текст попался, по трём причинам.

Первая: вы увидите, что именно сервис счёл важным. Список бывает неожиданным. У меня там в своё время оказалось несколько выводов обо мне, которые я не помню, чтобы сообщал, — они были собраны из контекста разговоров.

Вторая: вы поймёте, откуда берётся ощущение «она меня знает». Оно не мистическое, оно списочное. Заметки подкладываются в начало нового разговора, и модель отвечает, глядя в них.

Третья, практическая: устаревшая заметка портит ответы. Если полгода назад вы просили короткие ответы для конкретной задачи, а теперь вам нужны развёрнутые, старая инструкция будет тихо мешать.

Отдельно — про историю разговоров. Она хранится у сервиса и, в отличие от заметок, не подкладывается в новые чаты, но остаётся доступной вам и, в некоторых обстоятельствах, не только вам. Про то, что бывает с историей и почему кнопка «удалить» не всегда означает удаление, будет двадцать вторая глава.

Практическая привычка отсюда одна и простая: раз в несколько месяцев заходить в настройки, читать заметки, чистить лишнее. Занимает пять минут, а качество ответов от этого зависит сильнее, чем от смены сервиса.

Сегодня сделайте одно: найдите в настройках своего сервиса раздел про данные и память. Посмотрите, что там включено и что сервис успел о вас записать.

Пять минут — и вы будете знать про свой инструмент три вещи, которых не знает большинство: докуда он помнит мир, помнит ли он вас и что происходит с тем, что вы ему отдали.

Глава 6. Кто и как ей уже пользуется

Две картины, обе за 2026 год, обе из официальных источников.

Картина первая: семьдесят восемь процентов российских интернет-пользователей за год обращались к нейросетям, пятьдесят восемь процентов делают это раз в неделю и чаще. Это данные ВЦИОМ, опрос конца июня 2026 года, три тысячи двести человек.

Картина вторая: искусственный интеллект применяют четыре и восемь десятых процента российских организаций. Это расчёт Института статистических исследований и экономики знаний Высшей школы экономики на данных Росстата, опубликованный в марте 2026 года.

Семьдесят восемь и почти пять. Разрыв в шестнадцать раз.

Из этой пары чисел следует всё содержание главы: люди принесли ИИ в свою жизнь сами, поперёк работодателя, поперёк обучения, поперёк корпоративных регламентов. Никто им этого не поручал.

Кто эти люди на самом деле

Разложим семьдесят восемь процентов по группам, потому что среднее по стране всегда обманывает.

По возрасту, данные ВЦИОМ осени 2025 года. Восемнадцать-двадцать четыре года — восемьдесят один процент. Двадцать пять-тридцать четыре — семьдесят пять. Тридцать пять-сорок четыре — шестьдесят два. Сорок пять-пятьдесят девять — пятьдесят. Шестьдесят и старше — тридцать девять.

Заметьте, где стоит перелом. Не между молодыми и старыми, а между «до тридцати пяти» и «после». В группе, к которой относится большинство читателей этой книги, пользуется примерно каждый второй.

По цифровым навыкам, данные того же ВЦИОМ за июнь 2026 года: среди людей с минимальными навыками нейросети пробовали пятьдесят восемь процентов, с базовыми — семьдесят четыре, с продвинутыми — девяносто. Обратите внимание на первую цифру. Больше половины тех, кто считает себя технически беспомощным, уже заходили. Барьер входа ниже, чем страх входа.

По деньгам картина резче всего. Опрос SuperJob весной 2025 года, тысяча шестьсот экономически активных россиян из трёхсот шестидесяти семи городов: при доходе до пятидесяти тысяч рублей ИИ используют двадцать два процента, при доходе выше ста тысяч — пятьдесят четыре. Разница в два с половиной раза, и дело тут не в подписках — бесплатные сервисы доступны всем. Дело в характере работы: чем больше в ней текста, документов и решений, тем очевиднее польза.

И самая полезная для этой книги цифра, тоже из SuperJob: среди тех, кому за сорок пять, инструментом пользуются двадцать три процента. Против сорока двух в группе тридцать пять — сорок пять и шестидесяти шести среди тех, кому до тридцати пяти.

Тут придётся сделать оговорку, которую в новостях не делают. Данные ВЦИОМ и SuperJob для старших групп расходятся почти вдвое: пятьдесят процентов против двадцати трёх. Скорее всего, дело в выборках — ВЦИОМ опрашивает интернет-пользователей через онлайн-панель, SuperJob берёт экономически активных по всей стране. Правда где-то между, и я не знаю где. Когда вам в следующий раз покажут одну цифру про проникновение ИИ, помните, что у неё есть сестра, которая вдвое меньше.

Никто их не учил

Вернёмся к разрыву между семьюдесятью восемью процентами и пятью.

В июне 2026 года Институт статистических исследований ВШЭ опубликовал расчёт по данным Росстата. Навыками работы с искусственным интеллектом владеют тридцать семь с половиной процентов российских работников — из сорока шести и семи десятых миллиона

занятых. Базовый уровень, то есть чат-боты и генераторы картинок, — двадцать два и семь десятых процента. Средний — одиннадцать и семь. Продвинутый — три и две десятых.

А теперь цифра, ради которой я это привожу: только четыре и девять десятых процента работников говорят, что навыки ИИ нужны им по работе. В семь с половиной раз меньше, чем тех, кто ими владеет.

Прочитайте ещё раз. Люди научились тому, чего от них не требуют.

Это редчайшая ситуация. Обычно бывает наоборот: работодатель внедряет, обучает, требует, а сотрудники сопротивляются. Здесь сотрудники освоили инструмент сами и тихо им пользуются, а организация в массе своей не в курсе.

Подтверждение с другой стороны — опрос HeadHunter февраля 2026 года: ИИ в работе используют двадцать семь процентов россиян, а сорок пять процентов хотят, но не начали. Сорок пять. Это не сопротивление технологии. Это очередь у закрытой двери.

Отсюда практический вывод для читателя, который стесняется спросить у коллег. Скорее всего, половина вашего отдела уже пользуется и тоже стесняется. Разговор про нейросети на работе до сих пор происходит вполголоса, и не потому, что это запрещено, а потому, что непонятно, прилично ли.

Чем именно занимаются

Тут начинается самое интересное, потому что представление, сложившееся из новостей, расходится с реальностью почти во всём.

Осенью 2025 года OpenAI вместе с Национальным бюро экономических исследований разобрала около миллиона ста тысяч обезличенных сообщений — классификацией, без чтения людьми. Три крупнейшие категории: практические советы — около двадцати девяти процентов, поиск информации — около двадцати четырёх, работа с текстом — тоже около двадцати четырёх.

Не сочинение стихов. Не рисование картинок. Не написание кода. Люди спрашивают, как поступить, узнают факты и приводят в порядок тексты.

Российская картина похожа, но со своим акцентом. Яндекс разобрал обезличенные обращения к Алисе AI за период с сентября 2025 по январь 2026 года. Поиск информации — тридцать два и семь десятых процента. Медиа, то есть найти или разобрать картинку, аудио, видео, — восемнадцать. Общение и развлечение — четырнадцать, столько же советы и решение проблем. Работа с текстом — девять. Дальше совсем немного: покупки — три, математика и ИТ — три, кулинария — почти полтора.

И совсем маленькая деталь оттуда же, которая мне кажется важнее всех процентов. Люди старше сорока пяти чаще формулируют обращение как вопрос, а люди от восемнадцати до тридцати четырёх — как команду. Первые спрашивают, вторые поручают. В целом по обращениям: пятьдесят четыре и четыре десятых процента — вопросы, двадцать семь и три — команды, восемнадцать и три — утверждения.

Это не про вежливость. Это про то, кем человек считает эту программу. Собеседником, у которого спрашивают, или инструментом, которому поручают. Второе продуктивнее: поручение заставляет назвать результат, а вопрос позволяет этого не делать.

Сколько людей за это платят

Короткий раздел, который снимает лишнюю тревогу насчёт денег.

У ChatGPT в феврале 2026 года было девятьсот миллионов пользователей в неделю и пятьдесят миллионов платящих подписчиков. Это данные самой OpenAI. Посчитайте долю: около пяти с половиной процентов.

Девятнадцать из двадцати не платят ничего.

В России картина своя. По данным платёжного сервиса, обработавшего платежи с марта по август 2026 года, семьдесят четыре и три десятых процента платящих оплачивают ровно один сервис, а пятнадцать процентов пользователей дают пятьдесят семь процентов всего оборота. То есть платит меньшинство, и внутри этого меньшинства платит по-крупному совсем узкая группа профессионалов.

Аудитория российских бесплатных сервисов при этом вполне взрослая: недельная аудитория Алисы AI в первом квартале 2026 года — двадцать четыре миллиона человек.

Вывод простой и приятный: начинать с денег не надо. Подписка имеет смысл, когда вы уже поняли, чего вам не хватает в бесплатном. Про то, что почём и что бесплатно доступно в России, — следующая глава.

Три года назад всё было иначе

Чтобы оценить скорость, вернёмся в апрель 2023 года. ВЦИОМ тогда провёл первый опрос про нейросети, и цифры выглядят как из другой эпохи.

Лично пробовал нейросеть — один процент россиян. Один. Что-то знали о них шестьдесят три процента, но хорошо знали лишь двенадцать. Видели результат работы — двадцать восемь процентов, в основном картинки.

От одного процента до семидесяти восьми за три года. Я не знаю другой технологии, которая прошла бы этот путь так быстро; телефоны, интернет и банковские карты осваивались десятилетиями.

Отсюда, кстати, следует важное для читателя, который считает, что отстал. Массовой эта тема стала позавчера. Люди, которые сейчас выглядят опытными, начали на год-полтора раньше вас — не на двадцать лет.

Есть и обратная сторона такой скорости. Организации за ней не успевают, и это видно по данным Росстата, обработанным Институтом статистических исследований ВШЭ. Среди российских компаний, которые уже применяют ИИ, шестьдесят шесть процентов используют распознавание изображений, пятьдесят — системы поддержки решений, тридцать три — обработку текста и речи. Обследование охватило более пятнадцати тысяч организаций.

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.