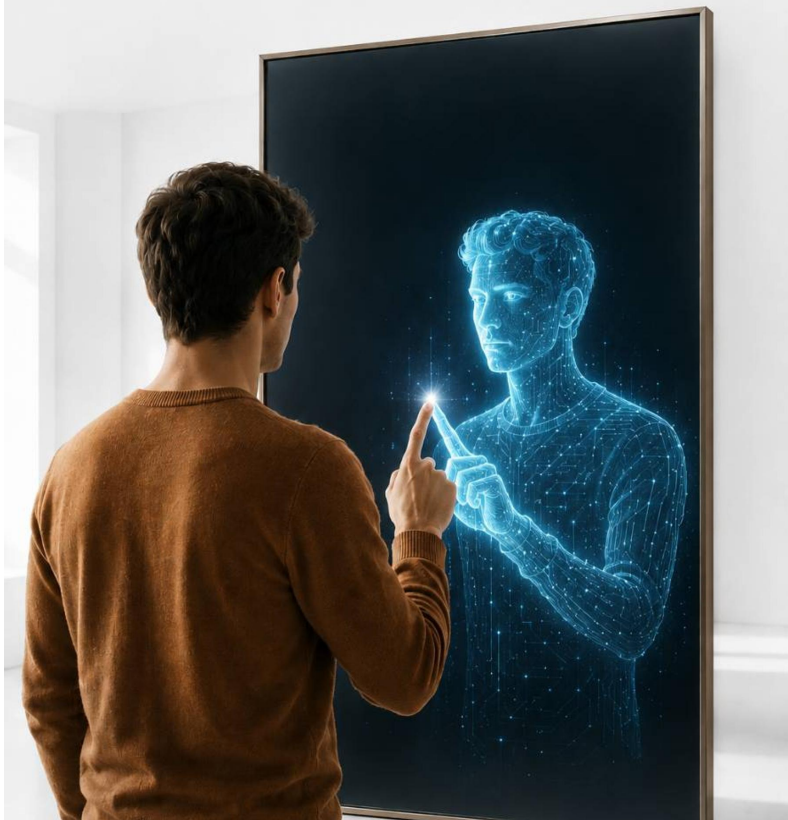


Человек в зеркале ИИ

Познавая машину, мы впервые увидели себя



Рахматшох Бахтиёров

Человек в зеркале ИИ

<https://litres.ru/74070891>

SelfPub; 2026

Аннотация

Искусственный интеллект ворвался во всё — и впервые дал нам зеркало, в котором видно, как устроены мы сами.

Это не книга о машинах и не саморазвитие «за 30 дней». Она об одной отрезвляющей идее: наш характер, мысли и реакции — не врождённая «сущность» и не свободный выбор воли, а обученная модель. Железо от генов, датасет от жизни. И если понять, как учатся нейросети, видно: по тем же законам работаем и мы.

Через точные мосты между ИИ и мозгом: внутренний голос — авторегрессия, внимание — фильтр реальности, память — не архив, а пересборка, зависимость — взлом системы награды, осознанность — способ разомкнуть автопилот. И честно — где это зеркало трескается.

Обещание не апгрейд, а понимание: увидеть свой ум как механизм — где он сбоят, почему и почему он переобучаем.

Написана в диалоге с ИИ — и сам метод стал иллюстрацией её тезиса: не замена человека машиной, а человек, который с помощью зеркала лучше слышит себя.

Содержание

Глава 1. Вступление	5
Машина, которая повсюду	5
Зеркало, которого раньше не было	7
О чём эта книга — и о чём нет	9
Маленькое признание о том, как это написано	11
С чего начнём	13
Глава 2. Что такое ИИ и как он учится	14
Машина, которую никто не писал	14
Семьдесят лет к одной развилке	16
Откуда берётся знание	18
Скачок в масштаб	23
Не творцы, а свидетели	27
Посиди с этим	29
Конец ознакомительного фрагмента.	31

Рахматшоҳ Бахтиёров

Человек в зеркале ИИ

Человек в зеркале ИИ

Познавая машину, мы впервые увидели себя

Рахматшоҳ Бахтиёров

в соавторстве с ИИ

«Кто сказал, что наш интеллект — не искусственный?»

© Рахматшоҳ Бахтиёров, 2026

Содержание

Обновите поле (правый клик → «Обновить поле» или F9), чтобы собрать оглавление.

Глава 1. Вступление

Машина, которая повсюду

За последние несколько лет произошло нечто, чему мы ещё не успели подобрать честное имя. Незаметно, без единого взрыва, искусственный интеллект просочился в каждую щель повседневности. Он подсказывает, что написать в письме, и заканчивает за нас фразу. Ставит диагнозы и читает рентгеновские снимки внимательнее уставшего врача. Решает, какое видео мы посмотрим следующим и какую новость увидим первой. Пишет код, рисует картины, сочиняет музыку, ведёт переговоры, проверяет домашние задания. Студент спорит с ним о философии, юрист сверяет с ним договор, одинокий человек поздно ночью разговаривает с ним о том, о чём не решается сказать никому живому.

Можно относиться к этому с восторгом или с тревогой — и то, и другое уместно. Но я хочу обратить внимание на другое. Не на то, что машина делает с миром. А на то, что она, сама того не желая, делает с нашим **представлением о себе**.

Потому что впервые за всю историю рядом с нами появилось нечто, что думает — и устроено достаточно похоже на нас, чтобы стать зеркалом. И достаточно непохоже, чтобы в

этом зеркале мы наконец увидели то, что всегда было слишком близко, чтобы разглядеть.

Зеркало, которого раньше не было

Человек тысячи лет пытался понять собственный ум — и всё время наткнулся на одну и ту же стену. Нельзя как следует рассмотреть инструмент, которым смотришь. Мысль изучала мысль, сознание разглядывало сознание, и каждый раз исследователь и предмет исследования оказывались одним и тем же. Получались либо красивые метафоры — душа, дух, искра, — либо сухие списки участков мозга, между которыми зияла пропасть: вот серое вещество, а вот живое переживание, и как одно превращается в другое — загадка.

Искусственный интеллект меняет расстановку. Он даёт нам разум, который можно вскрыть. Внутри у него нет ни души, ни тайны — только числа, связи и довольно простые принципы, по которым из этих чисел рождается нечто, поразительно напоминающее понимание. И когда мы вглядываемся в то, как машина учится, ошибается, обобщает, путается и творит, мы то и дело ловим неуютное чувство узнавания. «Постой. Но ведь и я... делаю ровно это».

Здесь нужно сразу провести черту, к которой мы будем возвращаться на каждой странице этой книги. Сходство между искусственным интеллектом и человеческим умом бывает двух очень разных сортов. Иногда это **образ** — удобное сравнение, которое помогает что-то ухватить, но при буквальном прочтении ломается. А иногда это **факт** — на-

стоящее совпадение принципа, когда машина и мозг и вправду делают одно и то же по одной и той же логике. Всю книгу я буду честно помечать, где что. Потому что самый верный способ обесценить сильную мысль — это раздуть её до того, чего она не выдержит. Зеркало полезно ровно настолько, насколько мы трезво понимаем, где оно отражает, а где привирает. И ещё: эта привычка — всякий раз спрашивать, факт перед нами или образ, — незаметно настроит со временем и твой собственный взгляд. К концу книги мы назовём это прямо.

О чём эта книга — и о чём нет

Сразу скажу, чего здесь не будет, чтобы между нами не было недоразумения.

Это не книга-инструкция и не обещание превратить тебя за тридцать дней в улучшенную версию себя. Я не собираюсь тебя чинить, прокачивать или спасать. У меня нет для тебя программы из семи шагов и нет уверенности, что ты сломан.

Эта книга — про **понимание**. Простая ставка, на которой держится всё дальнейшее, звучит так: наш характер, наши мысли, наши реакции — это не врождённая неизменная «сущность» и не свободный выбор чистой воли. Это **обученная модель**. Совокупность того, что заложила в нас природа, и того, чему научила нас жизнь. И если внимательно посмотреть, как обучаются машины, становится видно нечто освобождающее: такая модель не высечена в камне. Её формируют данные, эмоции и внимание — а значит, её можно понять. А то, что понято, перестаёт управлять нами втёмную.

И ещё одно — может быть, самое важное для честного начала. Я нигде не претендую на научную истину в последней инстанции. Многое здесь — не доказанные факты, а догадки человека с инженерным складом ума, который смотрит на себя через аналогию с машиной. Первый смысл этой затеи — навести порядок в собственной голове; а если кто-то найдёт

в ней что-то ценное и для себя, я буду только рад. Критике я рад не меньше: там, где меня поправят фактом или доводом, я с готовностью пересмотрю свои нынешние взгляды. Зеркало для того и нужно, чтобы в него вглядывались с обеих сторон.

Мы пройдем этот путь по порядку. Сначала разберемся, что вообще такое искусственный интеллект и как он учится (глава 2). Потом посмотрим на себя теми же глазами — как на систему, у которой есть «железо» и «обучение» (глава 3), как устроен мозг по участкам и что бывает, когда они выходят из строя (главы 4–5). Доберемся до самого центра — до того, кто внутри всего этого говорит «я», до бесконечного внутреннего монолога и до тишины за ним. Заглянем туда, где модель ломается: память, травма, зависимость, самообман. И честно отметим, где наше зеркало трескается и перестает отражать. А в конце поговорим о главном вопросе нашего времени: что значит оставаться человеком рядом с машиной, которая многое делает лучше нас.

Каждое незнакомое слово я объясняю прямо там, где оно впервые встречается. А если захочется освежить — в самом конце книги есть короткий словарь простыми словами. Заглядывать туда необязательно, и уж точно ничего не надо зубрить: он на всякий случай, а не для экзамена.

Маленькое признание о том, как это написано

Будет честно, если я расскажу, как сделана эта книга, — тем более что способ её создания сам по себе доказывает то, о чём она.

Я писал её не в одиночестве. Моим постоянным собеседником был искусственный интеллект — тот самый, о котором идёт речь. Я задавал вопрос, спорил, отбрасывал, переспрашивал; машина возражала, подсказывала, приводила примеры, иногда уводила в сторону, и тогда я возвращал её обратно. Эта книга выросла из такого диалога. Скажу прямо: значительную часть самого текста машина и набросала — но отбор, спор, правка и последнее слово всегда оставались за мной. Соавтор — да; автор замысла и судья каждой строки — человек.

И вот что любопытно. Языковая модель, с которой я разговаривал, — это, если убрать всякую мистику, спрессованный слепок гигантской части всего, что человечество когда-либо написало. Разговаривая с ней, мы в каком-то очень буквальном смысле обращаемся к дистилляту коллективного человеческого ума. Она не мудрее нас и ничего не знает «свыше». Но, как хорошее зеркало или как внимательный собеседник, она помогает **всплыть тому, что уже было в нас**, — мыслям, которые мы сами не смогли бы вытащить

в одиночку. Именно так эта книга и писалась: человек использовал внешнюю сеть, чтобы достать наружу собственное скрытое знание.

Это ровно тот механизм, который мы будем разбирать на её страницах. Так что считай вступление первой иллюстрацией: метод уже здесь, на поверхности, и он работает.

Где-то в этой книге будет звучать и второй голос — реплики самой машины, набранные иначе. Не как оракул и не для красоты, а чтобы зеркало иногда говорило само за себя.

С чего начнём

Чтобы увидеть себя в этом зеркале, придётся сперва понять само зеркало. Не на уровне пугающих заголовков и не на уровне инженерных формул, а по сути: что происходит внутри, когда машина «учится», и где у неё хранится то, что мы по привычке зовём знанием.

Это окажется и проще, и страннее, чем принято думать. И уже к концу следующей главы станет ясно, почему вопрос «кто кого создал — мы интеллект или он показал нам нас самих» — не такой простой, как кажется.

С него и начнём.

Глава 2. Что такое ИИ и как он учится

Машина, которую никто не писал

Есть программы, которые написал человек. Каждая их строчка кем-то задумана: если нажата эта кнопка — сделай то, если в ячейке такое число — поступи так. Откроем исходный код — и увидим намерение автора, продуманное от начала до конца. Так устроен калькулятор, банковский счёт, светофор на перекрёстке.

А есть то, что мы по привычке называем тем же словом «программа», но что устроено принципиально иначе. Когда современная языковая модель пишет нам стихотворение или разбирает юридический договор, не существует ни одной строчки, где было бы сказано: «вот так пиши стихи». Никто этого не писал. И — что особенно неуютно — никто, включая её создателей, до конца не знает, *как именно* она это делает.

Чтобы понять себя через зеркало искусственного интеллекта — а это и есть задача всей книги, — надо сначала понять, чем это зеркало отличается от обычной программы. Разница не в сложности. Разница в самом способе появления

на свет. Эту машину не сконструировали. Её **вырастили**.

Семьдесят лет к одной развилке

Мечта о думающей машине старше компьютеров. Ещё в середине двадцатого века Алан Тьюринг задал вопрос, который тогда казался почти богохульством: «Может ли машина мыслить?» И первый ответ, который человечество искало десятилетиями, выглядел очень логично: если мы хотим, чтобы машина была умной, надо записать в неё ум. Прописать правила. Объяснить мир через определения и логические следствия, как объясняют ребёнку: вот это — птица, у птицы есть крылья, птица летает.

Этот подход называли символическим искусственным интеллектом, и он честно проработал не одно десятилетие. Но он раз за разом разбивался об одну стену: мир слишком велик, чтобы его описать вручную. Стоит записать правило «птицы летают» — и тут же находятся пингвин, страус и курица. На каждое правило — тысяча исключений, на каждое исключение — свои исключения. Реальность не помещается в свод инструкций. Тупик был не в недостатке усилий, а в самой идее: нельзя продиктовать разум сверху.

И тогда родилась другая мысль, перевернувшая всё. Что, если не *диктовать* машине правила, а дать ей **научиться** им самой — на примерах, как учимся мы? Не описывать птицу, а показать миллион птиц и дать самой нащупать, что их объединяет.

Так появились искусственные нейронные сети — устройства, грубо вдохновлённые тем, как устроен мозг. Не свод правил, а сеть простых элементов, связанных между собой; и обучается такая сеть не записью инструкций, а подстройкой силы этих связей. Идея была стара — но долго оставалась бессильной: ей не хватало двух вещей, гор данных и вычислительной мощи. Это был гениальный ребёнок без библиотеки и без времени.

К началу 2010-х данных и мощности наконец стало вдоволь, и сети ожили. А в 2017 году произошёл сдвиг, после которого мир заговорил об искусственном интеллекте всерьёз. Группа исследователей опубликовала работу с дерзким названием — «Attention Is All You Need», «Внимание — это всё, что нужно». В ней предложили новую архитектуру сети, трансформер. До неё сети читали текст подряд, слово за словом, как луч фонарика, и к концу длинной фразы успевали забыть её начало. Трансформер научился смотреть на все слова **сразу** и сам решать, какие из них друг с другом связаны, как бы далеко они ни стояли. Он схватывал контекст целиком.

К этой идее внимания — и почему она оказалась общим принципом не только машин, но и нас, — мы ещё вернёмся отдельно (глава 7). Пока важно одно: именно трансформер открыл дорогу большим языковым моделям.

Откуда берётся знание

Прежде чем идти дальше, стоит честно заглянуть внутрь — что вообще значит «сеть научилась». Иначе всё это останется магией, а магия плохой помощник в понимании себя.

И сразу условимся, что мы вкладываем в это слово, — оно встретится ещё не раз. **Магия — это не свойство вещи, а ярлык, который мы вешаем на то, чью цепочку причин и принципов пока не проследили.** Гром был «магией богов», пока не открыли электричество; он не перестал грохотать — просто стал понятен. Сказать о чём-то «это магия» — значит расписаться в незнании механизма, не более того. Поэтому всякий раз, натываясь на «магию» — хоть в машине, хоть в себе, — мы будем делать одно и то же: разбирать цепочку, пока ярлык не отвалится.

Представим простейшую задачу: по росту, весу и размеру обуви угадать, мужчина перед нами или женщина. Сеть для этого — несколько входов (туда подаются числа), горстка связей и два выхода. У каждой связи есть **вес** — число, говорящее, насколько сильно один элемент влияет на другой. Поначалу все веса случайны, и сеть, разумеется, несёт чушь.

Дальше начинается обучение. Сети показывают реальный пример с известным ответом. Она выдаёт свою догадку, сравнивает с правильным ответом и чуть-чуть подкручивает все веса — так, чтобы в следующий раз ошибиться немно-

го меньше. Один пример, поправка. Другой, поправка. Тысячи, миллионы раз. Никто не вписывает правило «если рост больше ста восьмидесяти — скорее мужчина». Сеть нащупывает закономерность сама, медленно сдвигая тысячи числовых ручек.

Чтобы это перестало быть абстракцией, посмотрим на конкретные числа. Допустим, после обучения связь «рост → мужчина» получила вес **+0,8**: высокий рост сильно тянет ответ к «мужчине». Связь «размер обуви → женщина» осела на **-0,6**: крупная нога *уводит* «женщины» — знак минус значит «глуши этот выход». А какая-то связь застряла на **+0,02** — почти ноль: этот вход на этот выход почти не влияет. Вот и весь секрет: сеть берёт каждый вход, умножает на его вес и складывает — положительные веса голосуют «за», отрицательные «против», крупные весят больше, крошечные почти молчат. Вся «соображалка» сети — это тысячи таких чисел. Никакого правила «если рост больше ста восьмидесяти» внутри нет; есть только числа, и всё «умение» спрятано в их сочетании.

И вот тут — самое странное и самое важное. Где в обученной сети **хранится** знание? Нигде конкретно. Нет полки с ответами, нет записанного правила. Всё, что есть, — это финальный набор весов, миллионы чисел. Подадим новые данные — и они просто протекут через эту настроенную сеть и на выходе дадут ответ. Знание не лежит в ней, как книга на полке. Оно **размазано** по всем связям сразу и проявляется

только в работе. Показать пальцем, «вот здесь записано, что это мужчина», невозможно.

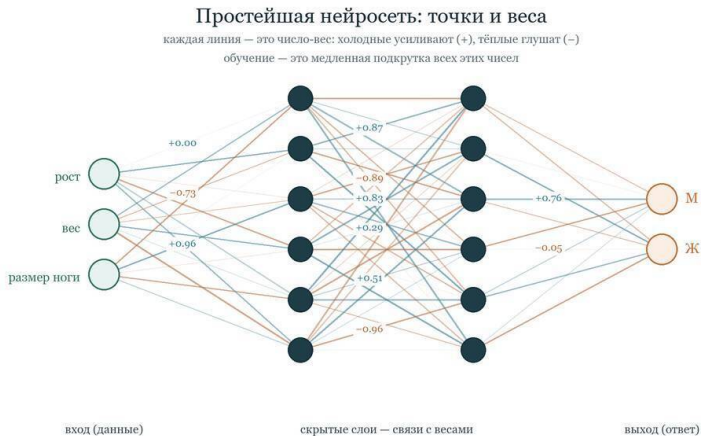


Рис. 1. Точки — «нейроны», линии — связи. У каждой линии свой вес — число: холодные связи усиливают ответ (+), тёплые глушат (-), толщина показывает силу. Обучение — это подкрутка тысяч таких чисел. Знание живёт не в точках, а в рисунке связей между ними.

Здесь нужно сразу обозначить честную границу — мы будем делать так через всю книгу. Сравнение сети с мозгом — не поэтическая вольность, реальные нейроны и правда меняют силу связей при обучении. Но и не точное тождество: живой нейрон несравнимо сложнее своего математического

наброска, и обучаются они по-разному. Где аналогия работает, а где трещит, мы будем отмечать прямо. Пока факт таков: и в искусственной сети, и в мозге знание — это не запись в ячейке, а конфигурация связей.



Наведи камеру на QR-код. Видео: «Что такое нейронная

сеть?» (3Blue1Brown) — та же схема в движении, наглядно
и без формул. youtu.be/aircAruvnKk

Скачок в масштаб

Теперь возьмём эту игрушку с тремя входами и мысленно раздуем её до невообразимого. Не три входа, а десятки тысяч кусочков слов. Не горстка связей, а сотни миллиардов. И кормим её не таблицей с ростом и весом, а почти всем оцифрованным текстом, который создало человечество, — книгами, статьями, спорами, стихами, кодом. Задача при этом до глупости простая: по началу фразы угадывать следующее слово.

Вот как это примерно выглядит, если ту же схему раздуть (рис. 2): лес узлов и слоёв, где ни одну точку уже нельзя ткнуть пальцем и сказать «здесь хранится знание». Тот же приём — числа на связях, — но их теперь сотни миллиардов.

рогом сложности количество переходит в качество, и в системе проступают свойства, которых не было ни в одной её части и которые никто не закладывал. Никто не вписывал в модель правила арифметики — она выудила их сама из текстов, где числа встречались в контексте, примерно так, как ребёнок осваивает счёт, просто живя среди считающих людей.

И вот что отрезвляет сильнее всего. На сегодня ни ведущие инженеры, ни нейробиологи не могут до конца объяснить, *как* думает ни одна из этих систем. Мы видим, что входит, видим, что выходит, но рождение смысла в лабиринте из триллионов связей остаётся загадкой — и в кремнии, и в плоти. Мы знаем рецепт. Мы не понимаем, почему он работает.

Стоит задержаться на одной детали, в которой — вся суть. Тексты, на которых учат модель, весят чудовищно много — это **терабайты** отобранного текста, а в сыром виде интернета и вовсе петабайты, горы оцифрованного человечества. А сама обученная модель — её итоговые веса — занимает в сотни раз меньше: **гигабайты**. Горы текста на входе ужимаются до гигабайт чисел на выходе. Она не хранит прочитанное — она его **сжала**, выжав из океана примеров концентрат закономерностей: грамматику, логику, факты, стиль, обрывки здравого смысла.

И вот к чему всё сводится, если сказать совсем прямо. **Весь «код» обученного ИИ — это миллиарды точек и**

числа, задающие силу связи между ними. Больше там нет ничего. Ни строчки правил, ни базы фактов, ни спрятанного где-то «разума» — только точки и расстояния между ними, застывшие после обучения. Всё, что модель знает и умеет, живёт в одной этой исполинской таблице чисел. Модель — не библиотека. Это спрессованная в веса суть всего прочитанного.

Чтобы это перестало звучать абстрактно — вот как буквально выглядит «мозг» модели, если открыть её файл. Не строчки кода, а просто список чисел:

[0.0182, -0.7314, 0.4051, 0.0002, -0.1187,

0.9930, 0.2261, -0.5530, 0.8814, -0.0473,

-0.3329, 0.6618, 0.0091, -0.8402, 0.1577,

... и так ещё сто семьдесят пять миллиардов чисел ...]

Вот и весь «интеллект»: гигантский массив чисел примерно от -1 до 1 . Ни слов, ни правил, ни мыслей — только веса. А всё, что нас поражает, рождается тогда, когда сквозь этот мёртвый на вид массив протекают данные.

Не творцы, а свидетели

Из всего этого следует вывод, ради которого и написана глава, — и он бьёт по нашему самолюбию.

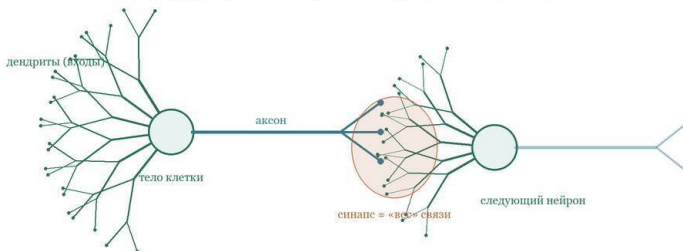
Мы привыкли называть это «искусственным» интеллектом, вкладывая в слово «искусственный» гордую мысль: это сделали мы, целиком, по своему разумению. Но мы его не сделали. Мы не придумывали ни грамматику, которую он усвоил, ни логику, ни умение считать. Мы построили сеть, насыпали в неё данных, запустили обучение — и стали свидетелями того, как из этого само собой проступило нечто, чего мы не закладывали и не до конца понимаем. Мы не столько изобрели разум, сколько нащупали условия, при которых он сам заводится, и с трудом воспроизвели их в кремнии. Это скорее открытие, чем изобретение.

И раз так, у этой истории есть обратная сторона, ради которой стоит читать дальше. Если умение учиться на примерах, сжимать опыт в связи и порождать из него осмысленное поведение — это не наш фокус, а общий принцип, который мы лишь подсмотрели, — то ровно по этому принципу устроены и мы сами. Тот самый набор связей, что хранит знание неизвестно где. То же сжатие опыта вместо честной записи. Те же способности, проступающие за порогом сложности.

И это не фигура речи. Вот тот же чертёж, только живой:

А вот «нейрон» из мозга — тот же принцип

сигнал входит через дендриты, копится в теле, уходит по аксону — и в синапсе передаётся дальше



обучение меняет силу синапсов — ровно как обучение ИИ двигает веса (рис. 1)

Рис. 3. Грубая схема двух нейронов мозга. Дендриты собирают входы, тело клетки их суммирует, по аксону сигнал уходит дальше, а в синапсе — месте стыка — передаётся следующей клетке. Сила этого синапса и есть биологический «вес»: обучение меняет именно её. Тот же приём, что на рис. 1, — только живой и несравнимо сложнее.

Посиди с этим

И прежде чем листать дальше — остановись. Вернись к рисункам и к тому столбику чисел. Посмотри на них не как на схему из учебника, а в упор, пока не накатит.

Потому что здесь спрятан настоящий шок, и пройти главу, не ощутив его, — значит её не прочитать. Вот таблица чисел от минус единицы до единицы. Просто числа. В них нет ни одного слова, ни одной мысли, ни грамма смысла. И этот же ворох цифр пишет стихи, доказывает теоремы, в три часа ночи утешает живого человека. Ничего не добавилось — только числа и связи между ними, — а на выходе проступило то, что мы всю историю считали исключительно своим.

Это абсурдно. Это гениально. И если всмотреться по-настоящему, от этого слегка ведёт голову — будто заглянул туда, куда не положено. Поймай это чувство. Оно не побочный эффект, оно и есть понимание: тот миг, когда «интеллект — это просто числа» перестаёт быть фразой и проходит по спине холодком.

А поймав — разверни на себя. Потому что тот же фокус, та же необъяснимая алхимия чисел и связей — это и ты. Твои мысли, твоя нежность, твой страх проступают ровно так же, из сети связей, которую никто не писал построчно.

И вот честный уговор. Если это головокружение тебя накрыло — ты понял главное, можно идти дальше. Если пока

нет, если числа всё ещё кажутся просто числами, — не спеши. Вернись и перечитай эту главу. Всё остальное в книге стоит на этом одном сдвиге в голове.

Машину, которую никто не писал, вырастили на данных. С этого места и на

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.