

КНИГА ДЛЯ ТЕХ, КТО НИЧЕГО НЕ ЗНАЕТ ПРО КОМПЬЮТЕРЫ

КАК УМЕСТИТЬ ВЕЛИКАНА В ЧЕМОДАН

Нейросети без формул

- Про то, как огромные умные машины делают маленькими и быстрыми

Ни одной формулы. Честное слово.



МЕНЬШЕ

• БЫСТРЕЕ •

УМНЕЕ

ЛУЧШЕ — ЧЕМ ЧТО?

Игорь Чепелев

**Как уместить великана в
чемодан. Нейросети без формул**

«Автор»

2026

Чепелев И.

Как уместить великана в чемодан. Нейросети без формул /
И. Чепелев — «Автор», 2026

Эта книга — про нейросети: огромные умные машины и способы сделать их меньше, легче и быстрее. Внутри нейросети нет ни фактов, ни словаря, ни ящичка с надписью «стихи» — только миллиарды чисел. Большинство из них можно округлить, многие — вообще убрать, и нейросеть почти ничего не заметит. Но некоторые числа трогать нельзя. Вся хитрость в том, чтобы понять, какие действительно важны. В 39 главах разберёмся, как уменьшить большую нейросеть, как вырастить маленькую вместо неё, как не заставлять её думать там, где можно просто посчитать, что при этом обычно ломается и — главное — как отличить настоящее улучшение от громкой надписи на коробке. Привычка проверять, стало ли что-то действительно лучше, пригодится далеко за пределами компьютеров. Для тех, кто ничего не знает про компьютеры и нейросети.

© Чепелев И., 2026

© Автор, 2026

Содержание

Введение. Белее, чем что?	6
Часть первая. Кто такой этот великан	7
Глава 1. Великан, который живёт в компьютере	7
Глава 2. Почему великан такой тяжёлый	8
Глава 3. Откуда взялись семь знаков	9
Глава 4. Почему великан долго думает (не то, что ты подумал)	10
Часть вторая. Делаем великана меньше	11
Глава 5. Линейка без миллиметров	11
Глава 6. Колокол и один баскетболист	12
Глава 7. Что трогать нельзя	13
Конец ознакомительного фрагмента.	14

Игорь Чепелев

Как уместить великана в чемодан. Нейросети без формул

Как уместить великана в чемодан

Про то, как огромные умные машины делают маленькими и быстрыми

Книга для тех, кто ничего не знает про компьютеры

Ни одной формулы. Честное слово.

Введение. Белее, чем что?

На пачке стирального порошка написано: «Теперь ещё белее!»

Прочитай это ещё раз и подумай минутку.

Белее — **чем что**? Белее, чем этот же порошок был раньше? Белее, чем у соседей? Белее, чем снег? Белее, чем варёное яйцо?

Ничего этого на пачке нет. Есть только слово «белее», а с чем сравнивают — не написали. И самое интересное: ты, скорее всего, даже не заметил подвоха. Наша голова терпеть не может пустых мест и сама вставляет туда что-нибудь подходящее.

А теперь то же самое, только про умные машины.

Люди делают такие машины — те, что отвечают на вопросы, пишут стихи и решают задачи. И постоянно пишут про них: «стала лучше», «работает быстрее», «почти не потеряла в качестве».

Лучше — чем что? Быстрее — чем что? Почти — это сколько?

Вот про это вся книга. Не только про то, как сделать машину меньше и быстрее — хотя про это тоже будет много всего интересного. А ещё и про очень полезную привычку: **всегда договаривать вторую половину сравнения.**

Что тебя ждёт.

Сначала мы разберёмся, из чего вообще сделана умная машина. Спойлер: не из проводов и не из микросхем — вернее, не только.

Потом узнаем, почему она такая тяжёлая и почему иногда думает медленно. Ответ окажется неожиданным.

Потом займёмся самым весёлым: будем делать её меньше. Разными способами. Их много, и каждый похож на какой-нибудь обычный домашний фокус — округлить, отрезать лишнее, научить младшего брата, не таскать вещи туда-сюда.

А в конце вернёмся к порошку и научимся не попадаться.

Одно обещание.

В этой книге не будет формул. Совсем. Не будет программирования. Не нужно ничего знать заранее.

Всё, что тебе понадобится, — уметь представлять картинки в голове. Например, вот прямо сейчас: представь чемодан. Готово? Значит, всё получится.

Часть первая. Кто такой этот великан

Глава 1. Великан, который живёт в компьютере

Представь очень большую таблицу. Таковую, как расчерченный на клетки лист, только огромную.

Не десять клеточек. Не тысячу. **Семь миллиардов клеточек.**

Это трудно вообразить, поэтому давай посчитаем по-другому. Если бы ты писал по одному числу в секунду, без сна, без еды и без выходных, тебе понадобилось бы больше двухсот лет. И это только чтобы записать. Один раз.

Вот эта таблица — и есть умная машина. Её настоящее имя — модель, но я буду называть её великаном: он огромный, тяжёлый и на самом деле довольно неповоротливый.

Что внутри каждой клеточки.

Одно число. Обычно маленькое — что-то вроде 0,37 или -0,08.

И всё. Больше ничего.

Чего внутри нет.

А вот это важнее. Внутри великана **нет** книжки с фактами. Нет словаря. Нет ящичка с надписью «стихи» и отдельного ящичка с надписью «математика».

Так думают почти все, кто слышит про умные машины впервые. Нет. Там только числа.

Тогда где же он хранит то, что знает?

Давай проверим. Спроси у такой машины: «Какая столица Франции?» Она ответит: «Париж».

Значит, она это знает. Значит, где-то это записано. Где?

Ответ странный: **нигде конкретно.** И одновременно везде понемножку.

Нет клеточки с надписью «Париж». Есть миллионы клеточек, и каждая чуть-чуть подталкивает ответ в нужную сторону. Одна подталкивает совсем незаметно. Другая тоже. Но их миллионы, и вместе они дают «Париж».

Это как хор. В хоре сто человек. Спроси, кто из них поёт мелодию, — никто по отдельности. Мелодия получается из всех вместе. Убери одного — ты даже не услышишь разницы. Убери восемьдесят — и песня развалится.

И вот отсюда начинается вся наша история.

Раз ни одна клеточка не главная сама по себе — значит, с этими числами можно обращаться не очень бережно. Округлить. Кое-где заменить на ноль. Кое-где выкинуть.

Великан этого почти не заметит. А чемодан станет намного легче.

Только надо знать, что можно трогать, а что нельзя. Об этом дальше вся книга.

Глава 2. Почему великан такой тяжёлый

Хорошо, семь миллиардов чисел. А сколько это места?

Тут надо разобраться с одной штукой, которая на первый взгляд скучная, а на самом деле в ней вся соль.

Сколько знаков после запятой?

Компьютер хранит число не как «0,7». Он хранит его подробно: **0,7003841**.

Семь знаков после запятой. У каждого из семи миллиардов чисел.

Зачем так подробно? Скоро узнаем — и это отличный вопрос. Пока просто запомним: очень подробно.

И сколько же это весит?

Одно такое подробное число занимает четыре байта. Байт — это мера, как грамм или сантиметр, только для компьютерной памяти.

Умножаем на семь миллиардов, получаем двадцать восемь гигабайт.

Само по себе это число мало о чём говорит, поэтому переведу. Двадцать восемь гигабайт — это примерно как:

семь полнометражных фильмов в хорошем качестве;

или целый сезон сериала;

или десять тысяч фотографий с телефона;

или столько, что в этот самый телефон оно просто не влезет. Вообще. Никак.

А ведь семь миллиардов — это **маленький** великан. Бывают в десять раз больше. И в пятьдесят.

Главный вопрос всей книги.

Вот теперь я задам вопрос, ради которого была вся эта глава.

Тебе правда нужны все семь знаков после запятой?

Подумай про чемодан. Ты собираешься в самолёт, и чемодан надо взвесить. На весах написано: **22,9847 килограмма**.

Скажи честно: тебе важно, что там 9847 в конце? Или тебе достаточно знать «двадцать три килограмма, значит, влезаю в норму»?

Достаточно двадцати трёх. Всё, что после запятой, — правда, но совершенно бесполезная правда. Ты не примешь ни одного решения по-другому, узнав про эти 9847.

А теперь представь семь миллиардов чемоданов.

Каждый взвешен до четвёртого знака. Каждый занимает место в памяти.

Сколько из них правда требуют такой точности?

Оказывается, почти ни один. Но некоторые — требуют, и очень настойчиво. Если их огрубить вместе со всеми, великан начнёт нести чушь.

Вся хитрость в том, чтобы отличить одни от других. Этим мы и будем заниматься.

Глава 3. Откуда взялись семь знаков

Погоди, скажешь ты. Если эти знаки не нужны, зачем их вообще записали?

Хороший вопрос. И ответ красивый.

Как великан учился.

Великан не родился умным. Его учили — очень долго, показывая огромное количество текстов.

Учёба выглядела так. Великану показывают кусочек текста и просят угадать следующее слово. Он угадывает. Обычно неправильно. Тогда все семь миллиардов чисел чуть-чуть подкручивают в ту сторону, где ответ был бы правильнее.

И так миллионы раз. Миллиарды.

Вот тут-то и нужны знаки.

Ключевое слово — «чуть-чуть». Каждая подкрутка **крошечная**. Не на 0,1 и даже не на 0,001. На что-то вроде одной стотысячной.

А теперь представь, что ты записываешь только два знака после запятой. У тебя число 0,37. Ты хочешь прибавить одну стотысячную.

$0,37 + 0,00001 = 0,37001$. Округляем до двух знаков... получаем 0,37.

Ничего не изменилось. Ты прибавил — и результат тот же самый. Как будто не прибавлял.

Если так делать всё время, великан вообще никогда ничему не научится. Он будет старательно подкручивать числа, а числа будут возвращаться на место.

Поэтому во время учёбы нужна большая точность. Обязательно.

А вот потом — не нужна.

Когда великан выучился, он больше ничего не подкручивает. Он просто считает: берёт твой вопрос и умножает на свои числа.

И вот здесь семь знаков — это ровно те самые граммы в аэропорту. Есть, честные, никому не нужные.

Вот и весь секрет.

Точность нужна, **пока учишься**. Когда выучился — можно округлять.

Это как черновик и чистовик. В черновике ты пишешь мелко, зачёркиваешь, считаешь в столбик, оставляешь пометки на полях. А в чистовик переписываешь только ответ.

Великан после учёбы — это чистовик. Черновик можно выбросить.

Глава 4. Почему великан долго думает (не то, что ты подумал)

Спроси кого угодно, почему большая умная машина отвечает медленно. Почти наверняка услышишь: «Ну она же большая, там много всего считается».

Логично. И неправильно.

Что происходит на самом деле.

Внутри компьютера есть две разные штуки, и их важно не путать.

Первая — **место, где считают**. Очень быстрое. Умножает числа с невероятной скоростью.

Вторая — **место, где хранят**. Гораздо больше, но добираться до него долго.

И вот великан начинает отвечать. Чтобы написать всего одно слово, ему нужно прогнать твой вопрос через **все** свои числа. Все семь миллиардов. Их надо достать из хранилища и принести туда, где считают.

Потом — следующее слово. И снова все семь миллиардов. Туда и обратно.

Ресторан, где повар быстрый, а склад далеко.

Представь ресторан. Повар готовит блюдо за три секунды — он гений, руки летают.

А продукты лежат на складе. В соседнем здании. И официант ходит за ними пешком.

Сколько блюд в час выдаст этот ресторан?

Столько, сколько успеет принести официант. Повар при этом почти всё время стоит и скучает.

Теперь скажи: если нанять второго повара, станет быстрее? А десятого? А сотого?

Нет. Ни на секунду. Потому что дело вообще не в поварах.

Вот и с великаном так же.

Его считалка почти всё время ждёт. Числа едут из хранилища, а она сидит без работы.

И теперь — самое интересное следствие, ради которого была эта глава.

Если мы сделаем каждое число **в четыре раза меньше** по размеру, что произойдёт?

Считать быстрее не станут — работы-то столько же. Но носить будут в четыре раза меньше!

Официант вместо четырёх ходок делает одну. Повар начинает работать почти без простоев.

Ответ ускоряется примерно вчетверо. И не потому, что мы упростили счёт, а потому, что перестали таскать лишнее.

Запомни этот фокус. Он объясняет почти всё, что будет дальше.

Часть вторая. Делаем великана меньше

Глава 5. Линейка без миллиметров

Пора сделать великана легче. Первый способ — самый простой, и он у тебя уже есть в голове.

Возьми линейку.

На обычной линейке есть миллиметровые деления. Мелкие, их много, глаза устают.

А теперь представь линейку, где остались только сантиметры. Деления крупные, их в десять раз меньше.

Можно ли ею мерить? Конечно. Просто теперь ты говоришь не «7 сантиметров 3 миллиметра», а «примерно 7 сантиметров».

Ты потерял точность. Зато линейка стала проще.

Вот и всё, что мы делаем с великаном.

У него числа записаны очень подробно. Мы говорим: хватит нам крупных делений.

Все числа великана лежат в каком-то промежутке — скажем, от -1 до $+1$. Мы режем этот промежуток на ступеньки. И каждое число заменяем **номером ближайшей ступеньки**.

Было 0,3172. Ближайшая ступенька — пятидесятая. Записываем: 50.

И всё! Вместо длинного подробного числа — короткое целое. Места надо в четыре раза меньше.

Когда великану понадобится это число обратно, он посмотрит на номер ступеньки и восстановит: получится 0,32.

Было 0,3172, стало 0,32. Ошибся меньше чем на процент. Великан этого даже не замечает.

Как это называется.

Квантизация. Слово страшное, звучит как из фантастики.

А означает оно ровно то, что мы только что сделали: округлить числа до ступенек и записать короче.

Так что если где-нибудь встретишь «квантизацию» — знай, речь про линейку без миллиметров.

Сколько ступенек оставить?

Тут есть выбор.

Оставим 256 ступенек — великан похудеет вдвое и почти ничего не заметит.

Оставим 16 ступенек — похудеет в четыре раза. Уже интереснее, но надо быть аккуратным.

Оставим 8 ступенек — очень мало. Тут великан начинает путаться.

Обычно останавливаются на шестнадцати. Это удачное место: и лёгкий, и умный.

Стоп, шестнадцать?!

Да, справедливое возмущение. Шестнадцать разных значений на все числа целого куска великана — как это может работать?

Об этом следующая глава. Там будет колокол.

Глава 6. Колокол и один баскетболист

Почему шестнадцать ступенек хватает? Из-за одной особенности, о которой стоит знать.

Числа лежат не как попало.

Если посмотреть на все числа великана и нарисовать, каких сколько, получится вот такая картинка: посередине огромная гора, а по краям — низенькие хвостики.

Это называется колокол. Он и правда похож на колокол, если нарисовать.

Что это значит по-человечески: **почти все числа маленькие и жмутся к нулю.** А больших — совсем чуть-чуть, единицы на тысячу.

И вот теперь понятно, почему шестнадцать ступенек справляются. Почти каждое число попадает в серединку, где ступеньки идут густо. Там всё в порядке.

А теперь про проблему.

Представь: ты измеряешь рост всех ребят в классе.

Все примерно от 140 до 160 сантиметров. Разброс маленький. Ты берёшь свою линейку с шестнадцатью делениями и растягиваешь их на этот промежуток — деления получаются мелкие, точность отличная.

И тут в класс заходит баскетболист. Двести десять сантиметров.

Что случилось? Теперь тебе надо мерить от 140 до 210. Промежуток стал в полтора раза шире. А делений всё те же шестнадцать.

Значит, **каждое деление стало в полтора раза грубее.** Для всех.

Один человек испортил измерение тридцати остальным. Причём сам он ни в чём не виноват — просто высокий.

Такие баскетболисты у великана есть всегда.

Их мало. Но они есть, и они очень важные — обычно как раз через них проходят самые сильные сигналы.

Так что просто выкинуть их нельзя. И оставить как есть тоже нельзя.

Что придумали.

Три хитрости, все понятные.

Первая: мерить по группам. Не растягивать одну линейку на весь класс, а разбить на компании по восемь человек. Тогда баскетболист испортит измерение только своим семерым соседям, а остальные будут измерены точно.

Вторая: заранее уменьшить баскетболиста. Перед измерением сказать: «Ты, высокий, присядь». Померили всех, а потом высокому вернули его рост обратно. Он же не изменился на самом деле — мы только на время его пригнули.

Третья: повернуть. Это самая хитрая и самая красивая. Если весь класс перемешать особым образом, то вместо одного очень высокого получится несколько просто высоковатых. Разброс станет поменьше, и линейка справится. А потом всё вернём на место.

Все три работают. Часто их применяют вместе.

Глава 7. Что трогать нельзя

У великана есть места, которые округлять **нельзя**. Их немного, и их полезно знать.

Вход и выход.

У великана есть начало — там твои слова превращаются в числа. И есть конец — там числа превращаются обратно в слова.

Это как переводчик на переговорах. Один переводит с русского на язык великана, другой обратно.

Если переводчик в середине разговора чуть ошибётся, остальные его поправят, разговор выправится. А если ошибётся тот, кто говорит **последнее слово**, — исправлять уже некому. Что сказал, то и услышали.

Поэтому вход и выход оставляют подробными. Стоит это немного, а спасает сильно.

Настройки громкости.

Внутри великана есть особые числа — их совсем мало, но они управляют громкостью всего остального.

Это как регулятор громкости у колонки. Сама колонка большая и сложная, а ручка одна. Крутанёшь чуть-чуть — изменится всё.

Округлять такие ручки нельзя. Экономии от них никакой (их же мало), а сломать можно всё.

Начало и конец очереди.

Великан состоит из слоёв — как бутерброд из множества слоёв. Числа проходят их по очереди, один за другим.

Оказалось, что первые и последние слои более обидчивые, а те, что в середине, — самые спокойные. Их можно округлять смелее всего.

Главное, что стоит понять.

Великан — не однородный кусок теста. Он разный в разных местах.

Одна из самых частых ошибок — обращаться со всеми его частями одинаково. «Округлим всё поровну» — звучит справедливо, а работает плохо.

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.