12+

Дмитрий Райтер

ИИ

Искусственный интеллект для начинающих

Как устроены нейросети
и как пользоваться ими с умом

Учебное пособие

Дмитрий Райтер

Искусственный интеллект для начинающих. Как устроены нейросети и как пользоваться ими с умом

«Издательские решения»

Райтер Д.

Искусственный интеллект для начинающих. Как устроены нейросети и как пользоваться ими с умом / Д. Райтер — «Издательские решения»,

Как работают нейросети и почему им нельзя верить на слово? Дмитрий Райтер объясняет основы искусственного интеллекта без сложной математики. История, российские разработки, мировой опыт, работа с запросами, проверка фактов и защита данных. В книге 54 главы, 18 практических работ с разборами и словарь. Для старшеклассников, родителей, педагогов и взрослых, которые начинают знакомство с ИИ.

© Райтер Д.
© Издательские решения

Содержание

Введение 6
Часть первая Что мы называем искусственным интеллектом 8
Глава 1 Когда программа кажется умной 8
Глава 2 Как возникла идея обучающейся машины 9
Глава 3 Почему одних правил оказалось недостаточно 10
Глава 4 От правил к обучению на примерах 11
Глава 5 Данные как материал для обучения 12
Глава 6 Ошибка и проверка качества 13
Глава 7 Как честно учить и проверять модель 14
Сравнение с простым способом 14
Глава 8 Нейронная сеть без сложных формул 16
Глава 9 Один искусственный нейрон на листе бумаги 17
Как заметить влияние одного веса 17
Часть вторая Как машина работает с языком 19
Глава 10 Текст превращается в числа 19
Глава 11 Что значит близость текстов 20
Похожая тема и противоположное действие 20
Глава 12 Откуда берётся ответ языковой модели 22
Глава 13 Вероятность без ложной уверенности 23
Двадцать прогнозов и осторожный вывод 23
Глава 14 Контекст и память 25
Глава 15 Поиск и генерация 26
Глава 16 Почему прочитанный файл ещё нужно проверить 27
Конец ознакомительного фрагмента. 28

Искусственный интеллект для начинающих Как устроены нейросети и как пользоваться ими с умом

Дмитрий Райтер

© Дмитрий Райтер, 2026

ISBN 978-5-0071-5190-0
Создано в интеллектуальной издательской системе Ridero

Введение

В российской школе, поликлинике, библиотеке и небольшой мастерской всё чаще возникает один и тот же вопрос: что можно поручить искусственному интеллекту? Одни ждут от него ответа на любую задачу. Другие опасаются, что он лишит людей работы и самостоятельности. Между ожиданиями и тревогами находится повседневная практика: составить понятное объявление, разобраться в незнакомом понятии, проверить ход решения, найти ошибку в программе, подготовиться к разговору со специалистом.

Эта книга начинается с российской действительности. Нам важны русский язык, доступные способы работы, защита сведений о людях и задачи, знакомые школьнику, родителю и учителю. Поэтому практические примеры опираются прежде всего на российские разработки: Алису AI, семейство YandexGPT, GigaChat, технологии Kandinsky, распознавание и синтез речи. При этом история и устройство искусственного интеллекта не ограничиваются одной страной. Наука развивается благодаря обмену идеями. Чтобы понимать современные системы, нужно знать и о мировых исследованиях, и о разных подходах к созданию моделей.

Нейросеть умеет убедительно говорить об ошибочном. Она может помочь разобраться в теме и одновременно предложить неверное объяснение. Поэтому главный результат знакомства с ИИ — не коллекция красивых запросов, а умение проверять. Читатель научится отделять сведения от предположений, замечать недостаток данных и выбирать задачи, в которых автоматическая помощь действительно полезна.

Пособие начинается с основных понятий и постепенно переходит к устройству моделей, российским сервисам и учебным проектам. Для чтения не нужны навыки программирования. Сложные понятия вводятся постепенно: сначала задача и наблюдение, затем объяснение и только после этого термин. Числовые примеры рассчитаны на обычную школьную арифметику. Там, где встречается условная модель, это прямо оговорено: учебный пример помогает увидеть принцип, но не воспроизводит всю работу промышленной системы.

Книга адресована прежде всего учащимся 8—11 классов и взрослым, которые начинают знакомство с ИИ. Она подходит для самостоятельного чтения, семейного обсуждения и подготовки занятий. Выбор материалов для конкретной образовательной программы остаётся за образовательной организацией в установленном порядке. Издание не заявляет о включении в федеральные перечни или о наличии ведомственного одобрения.

Практические задания можно выполнять без доступа к нейросети. Для этого в книге есть исходные тексты, условные ответы и ситуации для обсуждения. Если используется онлайн-сервис, взрослый предварительно проверяет его действующие условия, возрастные ограничения и правила образовательной организации. Возраст читателя книги и разрешённый возраст пользователя конкретного сервиса — разные вещи.

В примерах нет необходимости вводить фамилии одноклассников, домашние адреса, медицинские документы и фотографии людей. Любые учебные имена и ситуации вымышлены. Хорошая работа с технологией начинается раньше первого запроса: с решения, какие данные вообще не должны оказаться в чужой системе.

Как заниматься по этой книге

Читайте главы по порядку, если слова «модель», «данные» и «обучение» пока кажутся одинаково непонятными. Если основные принципы уже знакомы, можно перейти к российским сервисам, работе с запросами или практикуму. Разобранные расчёты и главы о школьных предметах позволяют закрепить основу на новых задачах. Словарь в конце поможет вернуться к нужному определению.

Не пытайтесь запомнить все названия продуктов. Они меняются быстрее, чем принципы работы. Полезнее научиться распознавать тип задачи: найти, объяснить, преобразовать, предсказать, создать или выполнить действие. Для каждого типа нужна своя проверка.

После чтения остановитесь на несколько минут. Сформулируйте главную мысль своими словами. Придумайте пример, которого нет в тексте. Если объяснение пока получается только дословно, вернитесь к разобранной ситуации. Пересказ своими словами показывает, что знание стало вашим.

Для занятий удобно завести журнал. Записывайте задачу, первоначальную попытку, полученную подсказку, проверку и собственный вывод. Не обязательно сохранять длинную переписку. Достаточно отметить, что именно изменилось в вашем понимании и на каком основании вы приняли ответ.

В заданиях встречаются вопросы с единственным проверяемым ответом и вопросы для обсуждения. В первом случае приведён ориентир для самопроверки. Во втором важнее обоснование. Два разных решения могут быть одинаково разумными, если их авторы понимают ограничения и последствия.

Часть первая Что мы называем искусственным интеллектом

Глава 1 Когда программа кажется умной

Ученик пишет в поисковой строке начало вопроса, а система предлагает продолжение. Телефон распознаёт произнесённую фразу. Приложение находит на фотографии нужный объект. Все эти действия выглядят привычными, хотя за ними могут стоять сложные модели. Искусственный интеллект давно встречается в повседневности и не сводится к одному чатботу.

Под ИИ обычно понимают область методов и систем, которые выполняют задачи, связанные с распознаванием, выводами, планированием, языком и другими интеллектуальными способностями. У этого понятия нет единственного бытового признака. Фраза «компьютер сделал что-то сложное» ещё не объясняет, какой метод использовался.

Обычный калькулятор уверенно умножает числа, но это само по себе не делает его обучающейся системой. Программа может работать по заранее описанным правилам. Например, электронный журнал считает средний балл по формуле. Разработчик задал порядок действий, и каждый шаг можно проследить.

Другую задачу трудно полностью описать правилами. Как отличить разборчивый почерк от почти такого же, но с другим смыслом? Как определить тему письма, если одно и то же слово используется в разных значениях? Здесь полезны методы, которые находят закономерности в примерах. Они не избавляют нас от правил, но позволяют решать задачи, где список условий оказался бы слишком длинным и хрупким.

Представим школьную библиотеку. Для выдачи книги по инвентарному номеру достаточно обычной базы данных. Для рекомендации следующей книги могут использоваться статистические методы и машинное обучение. Для составления описания выставки пригодна языковая модель. Одна библиотека, три задачи, три разных способа работы.

Слово «умный» часто служит рекламным определением. Полезнее задать конкретные вопросы. Что система получает на входе? Что выдаёт? На каких примерах её проверяли? Какую ошибку допускает? Можно ли отменить действие? Эти вопросы помогают обсуждать продукт без восторга и без страха.

Не следует приписывать программе человеческие переживания только потому, что она вежливо отвечает. Фраза «я рад помочь» может быть привычной формой диалога. Она не доказывает наличие радости. Для повседневной работы достаточно оценивать наблюдаемое поведение, а вопросы о сознании обсуждать отдельно.

Разберите три случая: будильник звонит в заданное время; программа распознаёт птицу на записи; чат составляет рассказ о поездке. В первом случае не требуется обучение на примерах. Во втором система распознаёт признаки. В третьем создаёт новый текст. Название «ИИ» полезно лишь тогда, когда после него мы способны объяснить различие.

Вопрос для самопроверки: почему автоматизация и искусственный интеллект не являются полными синонимами? Приведите пример автоматизации, в которой достаточно понятного правила.

Глава 2 Как возникла идея обучающейся машины

Попытки механизировать вычисления появились задолго до современных компьютеров. Однако вычисление по известному алгоритму и исследование машинного поведения — разные этапы истории. В XX веке развитие вычислительной техники позволило поставить вопрос: можно ли создавать программы, которые учатся и решают задачи, не сводящиеся к длинному перечню готовых ответов?

В 1950 году Алан Тьюринг опубликовал работу «Вычислительные машины и разум». Он предложил обсуждать способности машины через процедуру общения и наблюдаемые результаты. Позднее связанный с этой идеей подход получил широкую известность как тест Тьюринга. Успех в разговоре не равен доказательству сознания: собеседника можно убедить стилем, не обеспечив достоверности всех утверждений.

Название «искусственный интеллект» закрепилось за исследовательским направлением благодаря проекту Дартмутского семинара. Предложение было подготовлено в 1955 году, а сам семинар прошёл летом 1956 года. Эта разница полезна: дата документа и дата события не всегда совпадают. Исторический рассказ стоит проверять так же внимательно, как современную новость.

Ранние исследователи пробовали описывать рассуждение через символы и правила. Если известны посылки, программа выводит следствие. Такой подход хорошо работает там, где мир можно достаточно строго представить. Но повседневная речь полна пропусков, неоднозначности и знаний, которые человеку кажутся само собой разумеющимися.

Другой путь связан с обучением на примерах и моделями искусственных нейронов. Они не копируют мозг во всех подробностях. Это математические конструкции, вдохновлённые отдельными представлениями о передаче и преобразовании сигналов. Аналогия помогает начать объяснение, но становится ошибочной, если переносить на программу все свойства живого человека.

История не была непрерывным движением от успеха к успеху. Периоды больших обещаний сменялись разочарованием, когда вычислительных мощностей, данных или методов не хватало. Иногда такие спады называют «зимами искусственного интеллекта». Они напоминают: яркая демонстрация и надёжное решение массовой задачи — не одно и то же.

В России и ранее в СССР развивались исследования в области кибернетики, распознавания образов, математической логики, управления и вычислительных методов. Современные российские системы опираются на собственные инженерные школы и на международный научный опыт. Приписывать всю область одной компании или одному изобретению было бы историческим упрощением.

Составьте краткую хронологию без интернета: 1950 год — обсуждение машинного интеллекта через игру в имитацию; 1955 год — предложение исследовательского проекта; 1956 год — Дартмутский семинар. Затем объясните, почему история идеи не начинается в день появления популярного приложения.

Глава 3 Почему одних правил оказалось недостаточно

Представьте справочную школьного музея. На вопрос о времени открытия она должна отвечать по расписанию, а на вопрос о выдаче экспоната — направлять к ответственному сотруднику. Пока вопросов мало, удобно записать правила. Если в сообщении есть слово «расписание», показать часы; если есть слово «экспонат», вывести другое пояснение. Так можно построить полезную систему без языковой модели.

Трудности начинаются, когда посетитель спрашивает: «Успеем ли мы зайти после шестого урока?» В сообщении нет слова «расписание». Понадобится учесть время окончания урока, дорогу и режим музея. Следующий человек напишет «посмотреть выставку», а третий — «заглянуть на перемене». Правила разрастаются. Необязательно отказываться от них совсем; нужно понять, какую часть задачи они описывают удачно.

Исторически системы на знаниях и правилах позволяли фиксировать опыт специалистов в определённой области. Их называют экспертными системами. Если известны условия и правила перехода от них к выводам, программа может показать цепочку обоснования. Но знания приходится собирать, согласовывать и обновлять. Особенно трудно записать все исключения, которые специалист замечает благодаря опыту.

В задачах распознавания исследователи стали использовать обучаемые представления. Вместо перечня всех вариантов почерка модель подбирает зависимости по примерам. Этот подход выигрывает в разнообразии входов, но создаёт другой вопрос: как понять, чему именно она научилась? Точная запись правила уступает место проверке большого количества случаев и исследованию ошибок.

Важной вехой развития нейронных сетей стала опубликованная в 1986 году работа Дэвида Румельхарта и соавторов об обучении с обратным распространением ошибки. Её значение связано с подбором внутренних параметров многослойной сети. При этом история метода длиннее одной публикации: научные идеи нередко развиваются постепенно и в нескольких исследовательских группах.

Следующие успехи зависели не только от новых схем вычислений. Увеличивались наборы данных, совершенствовалось оборудование, появлялись удобные способы разработки и оценки. Нельзя объяснить появление современных помощников одной удачной формулой. Отдельная идея становится массовой технологией, когда сходятся несколько необходимых условий.

Сегодня правила и обучение часто работают вместе. Модель распознаёт намерение посетителя, поиск находит утверждённый документ, а обычная программа проверяет допустимое действие. Для выдачи книги лучше опираться на точную запись в системе, а не на свободно сгенерированное предположение. Разумная архитектура распределяет задачи между разными инструментами.

Попробуйте спроектировать музейную справочную на бумаге. Выделите пять вопросов, на которые достаточно точного правила, и три вопроса со свободными формулировками. Объясните, где полезна модель, а где её участие только усложнит проверку. Хорошее решение не обязано использовать ИИ в каждой части.

Глава 4 От правил к обучению на примерах

Допустим, нужно распределить письма школьной библиотеки по двум темам: продление выдачи и предложение новой книги. Можно написать правило: если встречается слово «продлить», отправить письмо в первую группу. Но посетитель напишет «можно оставить до пятницы», и правило не сработает. Чем разнообразнее язык, тем больше исключений приходится учитывать.

При машинном обучении системе показывают примеры писем с правильными метками. Она подбирает внутренние настройки так, чтобы по особенностям текста предсказывать тему. Такие настройки называют параметрами. Разработчик задаёт устройство модели, способ обучения и критерий качества, но не пишет отдельное правило для каждой будущей фразы.

Это не означает, что правила исчезли. До обучения всё равно нужно решить, какие категории существуют, что считать правильным ответом и как обрабатывать неоднозначные случаи. Если письмо одновременно просит продлить книгу и заказать новую, исходное деление уже не описывает ситуацию достаточно хорошо.

Обучение с учителем использует примеры с ответами или метками. При обучении без учителя модель ищет структуру без заранее выданной для каждого примера правильной категории. При обучении с подкреплением система получает сигнал о результате действия и подбирает поведение с учётом вознаграждения. Это крупные семейства подходов, внутри которых существует много вариантов.

Современные языковые модели часто сначала учатся на задачах, где целевой ответ формируется из самих данных: например, предсказывают продолжение текста. Такой подход называют самоконтролируемым обучением. Слово «само» здесь не значит, что программа сама придумала цель и независимо от людей решила развиваться.

Вернёмся к письмам. Если все учебные примеры о продлении начинаются со слова «Здравствуйте», модель может принять приветствие за важный признак темы. На новых письмах это приведёт к ошибкам. Обучение на примерах способно перенять случайную особенность выборки вместо нужной закономерности.

Поэтому часть данных оставляют для проверки. Проверочные примеры не должны незаметно попадать в обучение. Иначе получается как на контрольной, ответы к которой заранее разобрали и выдали за свидетельство самостоятельного знания. Высокий балл в такой ситуации мало говорит о работе с незнакомой задачей.

Учебное задание: придумайте четыре фразы с просьбой продлить книгу. В одной используйте слово «продлить», в другой обойдитесь без него, третью сделайте двусмысленной, четвёртую дополните предложением купить новое издание. Объясните, почему качество системы зависит и от модели, и от того, как люди определили задачу.

Глава 5 Данные как материал для обучения

Данные — это зафиксированные сведения: числа, тексты, изображения, звуковые записи, события и их признаки. Для человека десять фотографий одной чашки выглядят разнообразными. Для обучения распознаванию посуды этого может быть недостаточно: все снимки сделаны на одном столе и при одинаковом освещении.

Если модель увидела только белые чашки на тёмном фоне, она может плохо распознавать прозрачную кружку на светлой полке. Проблема не обязательно в количестве примеров. Важно, какие случаи представлены и насколько они похожи на будущие условия использования. Набор данных называют выборкой, а более широкий круг интересующих случаев — генеральной совокупностью.

Данные нужно готовить. Удаляют технические дубли, проверяют повреждённые файлы, согласуют обозначения и исправляют очевидные ошибки разметки. Однако очистка не должна превращаться в скрытое удаление всех неудобных примеров. Если исключить плохо освещённые фотографии, проверка может стать красивее, а реальная работа — хуже.

Разметка тоже требует решений. Два человека могут по-разному определить эмоциональную окраску фразы «Ну и отлично». В одном разговоре это искренняя радость, в другом — раздражение. Если контекст отсутствует, честной меткой иногда будет «недостаточно сведений». Принуждение к определённому ответу создаёт ложную точность ещё до обучения.

Представительность особенно важна для людей. Если речевую систему проверяли только на чётких записях взрослых дикторов, нельзя автоматически ожидать такого же результата на детской речи, диалектах и шумных разговорах. Ошибки могут распределяться неравномерно между группами пользователей.

Сведения о человеке не становятся безопасными только потому, что их много. Учебная выборка не должна собираться по принципу «скачаем всё, что найдём». Нужны законное основание, определённая цель и понимание прав на использование. Для школьной практики проще создавать небольшие вымышленные наборы, в которых нет реальных персональных данных.

Попробуйте составить десять вымышленных описаний погоды и отметить, пригоден ли день для наблюдения за облаками. Затем проверьте правило разметки: что вы будете делать ночью, при тумане или при сильном ветре? Неопределённость в инструкции быстро превратится в неопределённость модели.

Хороший набор данных сопровождается описанием: откуда он взят, какие случаи включает, чего в нём нет, когда его собирали и для чего его нельзя использовать. Такой паспорт помогает будущему пользователю понять пределы выводов. Файл без описания может содержать множество чисел и почти не содержать знания о том, что они означают.

Глава 6 Ошибка и проверка качества

Программа правильно ответила на девять вопросов из десяти. Это хороший результат? Пока неизвестно. Нужно понять, какие вопросы задавались, как часто встречается ошибка и что произойдёт из-за неверного ответа. Ошибка в названии цвета и ошибка в предупреждении о неисправности имеют разную цену.

Рассмотрим условный архив из ста писем. Девяносто пять относятся к обычной переписке и пять содержат срочную заявку. Если система всегда отвечает «обычное письмо», её доля верных ответов составит 95 процентов. При этом она не найдёт ни одной срочной заявки. Одна общая цифра скрыла провал по важной задаче.

Для классификации полезно различать два вопроса. Среди отмеченных системой срочных писем сколько действительно срочные? Это связано с точностью положительных срабатываний. Среди всех настоящих срочных писем сколько найдено? Это связано с полнотой. Термины можно уточнить позже; сначала важно увидеть разные виды ошибок.

Ложная тревога возникает, когда система объявляет проблему там, где её нет. Пропуск возникает, когда настоящую проблему не замечают. Настройка порога может менять соотношение этих ошибок. Более чувствительный фильтр найдёт больше подозрительных случаев, но может чаще отвлекать проверяющего.

У языкового помощника качество сложнее свести к одной величине. Ответ бывает грамотным, но неточным; полезным, но неполным; верным по сути, но неподходящим возрасту читателя. Поэтому заранее задают критерии: фактическая точность, понятность, соблюдение ограничений, наличие необходимых сведений и отсутствие лишних утверждений.

Проверка должна включать обычные и трудные случаи. Если помощник составляет объявления, предложите ему неполные данные, противоречивые даты и просьбу придумать отсутствующий адрес. Хороший результат иногда состоит в уточняющем вопросе или явном указании на пробел. Умение остановиться может быть полезнее длинного уверенного текста.

Не выбирайте только удачные ответы. Если из десяти попыток показали одну, наблюдатель не узнает, насколько результат воспроизводим. Сохраняйте условия проверки: дату, версию сервиса, запрос, исходные сведения и способ оценки. Без этого сравнение легко превращается в соревнование впечатлений.

Задание: в архиве двадцать писем, из них четыре срочных. Система пометила пять писем, среди которых три действительно срочные. Сколько было ложных тревог и пропусков? Ответ разберите словами, затем проверьте в разделе самопроверки. Подумайте, какую ошибку было бы опаснее допустить в вашей задаче и почему.

Глава 7 Как честно учить и проверять модель

Вообразим учебный набор из девяноста вымышленных описаний растений. Для каждого заранее указана категория. Участники хотят подобрать простую систему и узнать, сможет ли она распределять незнакомые описания. Чтобы проверка имела смысл, примеры делят по назначению до подбора решения.

Одна часть используется для обучения: по ней настраиваются параметры. Другая помогает выбирать настройки и варианты системы. Третья сохраняется для итоговой проверки. Для упражнения можно выделить 60, 15 и 15 описаний соответственно. Это не универсальная правильная пропорция, а удобный числовой пример.

Если после каждого неудачного результата на последних пятнадцати примерах менять решение, эти примеры перестают быть независимой итоговой проверкой. Вы постепенно подгоняете систему под них. Внешне файл всё ещё называется «тест», но его роль изменилась. Название папки не создаёт независимость.

Есть и более незаметная утечка. В наборе оказались почти одинаковые описания одного растения с небольшими изменениями слов. Одно попало в обучение, другое — в тест. Результат может выглядеть высоким потому, что система уже видела почти тот же случай. При разделении иногда нужно учитывать общность объектов и происхождение примеров.

Подобное случается и без обучения новой модели. Если человек двадцать раз меняет запрос, ориентируясь на один пример, а затем показывает успех только на нём, мы мало знаем о пригодности запроса для новых задач. Нужен набор случаев, которые не использовались при подборе формулировки.

Зафиксируйте исходный способ до начала испытания. Сначала простое правило или работа без ИИ, затем предложенный помощник. Иначе невозможно установить, стало ли лучше. Сложная система должна приносить проверяемое преимущество, а не побеждать только за счёт отсутствия разумного сравнения.

Результаты показывают полностью. Укажите число попыток, типы ошибок и случаи, в которых система просила уточнение. Если два ответа из пятнадцати непригодны, не заменяйте их незаметно повторной генерацией. Повторные попытки можно исследовать отдельно, сохранив описание процедуры.

Задание: подготовьте шесть коротких учебных объявлений для настройки и ещё четыре для окончательной проверки. Одно проверочное объявление сделайте неполным, другое противоречивым. Объясните, почему нельзя заранее подстроить запрос под все десять и затем назвать четыре из них незнакомыми.

Сравнение с простым способом

До создания помощника участники вручную проверяли объявление по трём вопросам: есть ли дата, место и условие участия. На двенадцати учебных текстах они нашли все заранее заложенные пропуски за восемь минут. Новый помощник предложил ответы за минуту, но проверка добавленных им сведений заняла ещё десять минут. Можно ли уверенно говорить об экономии времени?

Нужно учитывать весь процесс. В этом условном опыте автоматическая подготовка оказалась быстрее, а проверка увеличила общую длительность до одиннадцати минут. Это не доказывает бесполезности инструмента во всех задачах. Но заявлять выигрыш на данном процессе нельзя без изменения процедуры или новых сопоставимых испытаний.

Если помощник заодно обнаружил противоречие, которого не было в ручной памятке, это отдельная польза. Её нужно описать прямо и решить, стоит ли она дополнительного времени. Разные показатели не обязательно сводятся к одной победной цифре.

Глава 8 Нейронная сеть без сложных формул

Представим простую систему, которая оценивает, хватит ли времени на прогулку до начала кружка. На входе три числа: оставшиеся минуты, длительность дороги и запас на непредвиденную задержку. Система должна связать входные сведения с результатом. Настоящая нейронная сеть может иметь гораздо больше входов и внутренних преобразований, но принцип удобно обсуждать на небольшом примере.

Искусственный нейрон получает числовые сигналы, учитывает их с разными коэффициентами и преобразует результат. Коэффициенты называют весами. Помимо весов могут использоваться смещения и нелинейные преобразования. Вместе они позволяют строить более сложные зависимости, чем простое сложение.

Нейроны объединяют в слои. Один слой преобразует представление данных, следующий работает с полученным результатом. В задачах изображений ранние этапы могут выделять простые признаки, а последующие сочетать их. Не следует понимать это как обязательную понятную человеку лестницу «линия — глаз — лицо» в любой сети. Реальные представления бывают распределёнными и трудно объяснимыми.

При обучении модель делает предсказание, его сравнивают с целевым результатом и вычисляют величину ошибки. Затем параметры меняют так, чтобы на следующих шагах ошибка уменьшалась. Широко применяются методы градиентной оптимизации и обратного распространения ошибки. Для первого знакомства достаточно понимать направление: система многократно корректирует числовые настройки по обучающему сигналу.

Отдельный пример не всегда улучшает работу на всех остальных. Если слишком усердно подгонять модель под учебную выборку, она может запомнить её особенности и хуже работать на новых случаях. Это называют переобучением. Аналогия со школьником проста: можно выучить ответы конкретного сборника, не освоив способ решения.

В работе модели различают обучение и применение. Обучение обычно требует существенно больше ресурсов. При применении уже настроенная модель обрабатывает новый ввод. Продолжение разговора может менять текущий ответ за счёт контекста, но это не обязательно означает изменение весов модели после каждой реплики.

Название «нейронная сеть» не даёт оснований считать её цифровым мозгом в полном биологическом смысле. Полезность математической модели определяется результатами и проверкой, а не похожестью названия на части организма. Не нужно выбирать между мистикой и пренебрежением: перед нами мощный вычислительный инструмент со своими ограничениями.

Объясните младшему товарищу различие между весами и текущим запросом. Веса — это настройки, полученные при обучении. Запрос — сведения, поступившие сейчас. Если объяснение получилось без слов «она просто всё понимает», вы сделали важный шаг к точности.

Глава 9 Один искусственный нейрон на листе бумаги

Слово «веса» становится понятнее, если выполнить небольшой расчёт. Рассмотрим условную систему, которая выбирает учебные статьи для кружка. У каждой статьи есть два признака: присутствует ли слово «робот» и присутствует ли слово «выставка». Для наличия признака запишем 1, для отсутствия — 0. Это очень грубое описание текста, созданное только для упражнения.

Пусть вес первого признака равен 2, второго — 1. К сумме добавляется смещение, равное минус 1,5. Умножаем каждый признак на его вес, складываем и учитываем смещение. Если результат больше нуля, статья попадает в подборку. Иначе она не выбирается.

В статье есть оба слова. Получаем 2 плюс 1 минус 1,5, то есть 1,5. Условие выполнено. В статье есть только «робот»: 2 минус 1,5 равно 0,5, статья тоже выбирается. Если есть только «выставка», результат равен минус 0,5, поэтому статья не проходит. Без обоих слов результат равен минус 1,5.

Заметьте, слово «робот» в этой игрушечной системе влияет сильнее. Именно так удобно начать понимать вес: это коэффициент, меняющий вклад входного признака. Смещение сдвигает условие решения. Но у настоящей модели признаки могут быть сложными числовыми представлениями, а вычисления — состоять из множества связанных преобразований.

Теперь возникает содержательная ошибка. Статья «Робот на упаковке игрушки» проходит, хотя кружку нужны материалы о технике. А текст «Автоматическая машина на школьной выставке» не проходит из-за отсутствия выбранного слова. Ошибка связана не с арифметикой, а с бедным описанием содержания. Правильно посчитать по плохим признакам недостаточно.

Изменение весов тоже не решит всех проблем. Если два разных по смыслу текста имеют одинаковые входные признаки, эта система не сможет их различить. У неё просто нет информации, которая объясняет различие. В таких случаях нужны другие признаки, более подходящие данные или новая постановка задачи.

Пороговая функция в примере выбрана для наглядности. В обучаемых нейронных сетях часто используют другие функции преобразования, позволяющие эффективно настраивать параметры. Не переносите механически наш способ ручного решения на любую современную архитектуру. Учебная модель показывает отдельную идею, а не уменьшенную копию большого чат-помощника.

Составьте четыре собственных заголовка и вычислите решения. Затем измените вес второго слова на 2. Какие статьи стали проходить? Укажите хотя бы один случай, где новый результат всё равно не соответствует цели кружка. Так видно различие между настройкой числа и улучшением смысла задачи.

Как заметить влияние одного веса

Пусть статья о выставке не проходит, а по решению руководителя кружка должна проходить. У неё признаки 0 и 1. При прежнем втором весе 1 получается минус 0,5. Если заменить его на 1,6, результат станет 0,1 и решение изменится. Это показывает направление ручной корректировки, но ещё не описывает полноценный алгоритм обучения.

Проверьте остальные три комбинации. Статья с обоими признаками получает 2,1, только с первым — 0,5, без признаков — минус 1,5. Изменился не один сохранённый ответ, а правило обработки всех новых входов. Поэтому настройка параметра требует проверки других случаев, а не только того, ради которого внесли исправление.

Теперь предположим, что часть материалов о выставках не подходит кружку. Увеличенный вес будет пропускать и их. Задачу нельзя решить бесконечным подбором одного коэффи-

циента: нужен дополнительный признак, различающий подходящие и неподходящие тексты. Так маленький расчёт показывает, почему устройство данных иногда важнее размера модели.

Часть вторая Как машина работает с языком

Глава 10 Текст превращается в числа

Человек видит в предложении смысл, интонацию и намерение собеседника. Вычислительная система получает последовательность кодов и выполняет операции над числовыми представлениями. Между набранной фразой и ответом находится несколько преобразований. Одно из первых — разбиение текста на токены.

Токен может совпадать со словом, частью слова, знаком препинания или другой единицей. Правило зависит от токенизатора. Поэтому сто слов не обязательно равны ста токенам. Русская и английская фразы одинаковой длины могут занимать разный объём. Точные ограничения сервиса нужно проверять в его документации, а не выводить из универсального коэффициента.

После разбиения токены переводятся в числовые представления. Удобно представить их как точки или направления в пространстве с большим числом измерений. Такие представления могут отражать закономерности употребления, но их не следует воспринимать как словарные карточки с единственным значением каждого слова.

Фраза «ключ лежит на столе» отличается от фразы «ключ бьёт из скалы». Контекст помогает системе выбрать подходящие связи. Именно поэтому короткий вопрос без пояснений иногда получает формально грамотный ответ совсем о другом предмете. Пользователь знает, о чём думал, а программа видит только доступный ей ввод.

В современных языковых моделях важную роль играет механизм внимания. Он позволяет при обработке учитывать связи между частями последовательности. Архитектура трансформера была представлена в работе «Attention Is All You Need» в 2017 году. Это один из поворотных этапов развития, но не единственная причина последующего успеха: важны также данные, вычисления, обучение и организация применения.

Название «внимание» — техническая аналогия. Оно не доказывает, что программа сосредоточивается и отвлекается так же, как человек. Вес связи в модели и человеческое переживание значимости — разные уровни описания. Слишком буквальные аналогии мешают заметить настоящие причины ошибок.

Практический вывод прост. Если нужно обсудить школьный проект о роднике, напишите это в запросе. Не полагайтесь на то, что слово «ключ» будет понято само собой. Укажите цель, аудиторию и исходные сведения. Дополнительный контекст не гарантирует безошибочности, но уменьшает ненужную неоднозначность.

Задание без компьютера: придумайте по два предложения со словами «коса», «лук» и «язык» в разных значениях. Затем сформулируйте вопрос, на который нельзя ответить однозначно без предыдущего предложения. Так вы увидите, какую работу контекст выполняет ещё до проверки фактов.

Глава 11 Что значит близость текстов

Два сообщения могут сообщать одно и то же разными словами. «Верните книгу до пятницы» и «Последний день возврата — четверг» похожи по теме, но не полностью совпадают по условиям. Поиск должен замечать близость и одновременно сохранять различия, влияющие на действие читателя.

Для работы с похожестью используются векторные представления. Начнём с простого вымышленного пространства из двух признаков: сколько в тексте сведений о книгах и сколько — о спортивных занятиях. Условное описание библиотеки имеет координаты 4 и 0, описание стадиона — 0 и 4, а рассказ о чтении спортивной энциклопедии — 3 и 2. Настоящие представления устроены значительно сложнее и обычно не имеют таких подписанных человеку осей.

По координатам можно вычислять разные меры близости. Если рядом в этом пространстве находятся тексты о книгах, поиск сможет находить тематически родственные документы без полного совпадения слов. Но выбранная мера и качество представления влияют на результат. Близость не является доказательством истинности, а расстояние не заменяет чтение.

Представим базу правил библиотеки. Запрос «Как продлить книгу?» должен привести к документу о продлении. Однако рядом может оказаться старая версия, в которой указан другой срок. Поэтому кроме смысловой близости нужны признаки документа: дата, статус, название учреждения и область применения. Удачный поиск сочетает несколько оснований.

Полезно различать поиск по точному совпадению и поиск по смыслу. Если нужен номер приказа, точная строка часто надёжнее широкого смыслового соответствия. Если пользователь не знает официального термина, смысловой поиск помогает начать. Выбор зависит от того, какую неопределённость требуется устранить.

Векторное представление иногда описывают словами «цифровой смысл». Это удобная краткость, но её легко понять слишком буквально. Представление сохраняет те закономерности, которые модель научилась выделять. Оно не обязано включать все важные человеку различия. Два предложения с отрицанием могут оказаться близкими по теме и противоположными по действию.

Поэтому система поиска по документам должна не только находить, но и показывать основание. Пользователю полезно видеть найденный фрагмент в контексте и дату документа. Если ответа нет, корректнее сообщить об этом, чем составить красивое продолжение по соседним темам.

Упражнение: напишите три вопроса о выдаче книги, используя разные слова. Затем добавьте вопрос, в котором меняется срок или действует исключение. Объясните, какие ответы можно объединить, а какие обязательно различать. Проверяйте условия, даже когда тема кажется одинаковой.

Похожая тема и противоположное действие

Рассмотрим два документа: «Запись на встречу обязательна» и «Запись на встречу не обязательна». Они почти одинаковы по словам и относятся к одной ситуации. Для смыслового поиска их соседство разумно. Для ответа посетителю различие решающее: в одном случае нужно записаться, в другом такого требования нет.

Поэтому полезно разделить работу на два шага. Сначала найти материалы по теме встречи. Затем определить точное правило из действующего документа. Если система объединяет обе фразы в «запись обычно нужна», она создаёт новое условие, которого нет ни в одном источнике. Усреднение противоречивых правил не устраняет противоречие.

Тот же принцип помогает читать научные и исторические материалы. Два текста могут обсуждать одну гипотезу, но один её поддерживает, а другой оспаривает. Близость темы нужна для поиска дискуссии, но позицию каждого автора придётся устанавливать отдельно.

Глава 12 Откуда берётся ответ языковой модели

В упрощённом описании языковая модель последовательно выбирает продолжение на основе вероятностей, рассчитанных по доступному контексту. Она учитывает не только последние слова, но и другие части входной последовательности в пределах своих возможностей. На выход влияют параметры модели, инструкции, история диалога и настройки генерации.

Это описание не означает, что ответ всегда сводится к случайному угадыванию одного слова. Обученные представления позволяют выполнять сложные преобразования: объяснять, переводить, обобщать и строить рассуждения. Но способность получать полезный текст не создаёт автоматической гарантии истинности каждого утверждения.

Если попросить продолжить фразу «Столица России — », ожидаемое продолжение обычно очевидно. Если попросить назвать автора редкой статьи с неясным заголовком, модель тоже может продолжить уверенно. Снаружи ответы выглядят одинаково гладкими, хотя основания для доверия различаются. Стиль не является шкалой достоверности.

Один запрос может давать разные ответы. Иногда это связано со случайным выбором продолжения, иногда — с обновлениями системы или скрытыми различиями условий. Настройки, повышающие разнообразие, полезны при поиске идей. В задаче точного извлечения фактов они могут создавать лишнюю вариативность. Даже более сдержанная генерация не превращает модель в непогрешимый справочник.

Модель может опираться на сведения, усвоенные при обучении, на предоставленные документы или на результаты подключённого поиска. Это разные источники. Если доступ к поиску не включён, просьба «посмотри сегодняшнее расписание» сама по себе не создаст интернет-соединение. Система либо использует доступный инструмент, либо остаётся в пределах имеющихся данных.

Не путайте правдоподобную ссылку с открытым источником. Модель способна составить адрес, похожий на настоящий, и назвать несуществующую публикацию. Чтобы проверить ссылку, её открывают, смотрят автора, дату и содержание. Наличие списка литературы в ответе не заменяет чтения хотя бы тех мест, которые подтверждают важные утверждения.

Полезный запрос: «Если исходных сведений недостаточно, перечисли пробелы. Не придумывай имена, даты и ссылки». Такая инструкция помогает обозначить ожидания, но не отменяет проверки. Сервис может нарушить ограничение. Ответственность пользователя состоит в том, чтобы заметить это до дальнейшего использования.

Проведите мысленный опыт. Даны две фразы: «встреча состоится в 15:00» и «время встречи ещё не согласовано». Попросили составить окончательное приглашение. Надёжный помощник должен заметить конфликт, а не выбрать красивую формулировку и скрыть неопределённость. Качество начинается с верного отношения к исходным данным.

Глава 13 Вероятность без ложной уверенности

Предположим, система оценила вероятность дождя как 70 процентов. Это не означает, что дождь займёт 70 процентов дня или выпадет на 70 процентов территории. Значение зависит от того, какое событие и для каких условий прогнозируется. Без определения события число легко получить и трудно правильно использовать.

В условном учебном опыте вы собрали сто сопоставимых прогнозов, каждому из которых приписывалась вероятность 70 процентов. Если дождь происходил примерно в семидесяти случаях, прогнозы по этому признаку выглядят согласованными с наблюдениями. Так можно начать понимать калибровку. Но для настоящей оценки нужны подходящие данные, достаточный объём и ясные условия.

Одно событие не проверяет качество вероятностного прогноза целиком. Дождь мог не пойти и при 70 процентах. Это не доказывает, что прогноз был бессмысленным. И наоборот, один удачный исход не делает систему надёжной. Оценка требует серии случаев и правильного способа сравнения.

В диалоге с языковой моделью есть другая трудность. Просьба «оцени уверенность в процентах» может дать число, которое выглядит научно, но не получено из проверенной процедуры. Не принимайте его за измеренную вероятность истинности ответа. Словесное признание сомнений полезно, однако его тоже нужно сопоставлять с основаниями.

Вероятность следующего токена и достоверность всего сообщения — разные величины. Привычная фраза может иметь высокую вероятность продолжения, хотя относится к неверной предпосылке. Последовательность «как известно, все…» способна звучать естественно и содержать ложное обобщение. Уверенный языковой рисунок не заменяет проверки факта.

Решение зависит не только от вероятности, но и от последствий. Если возможен небольшой дождь, можно взять зонт. Если существует даже сравнительно небольшая вероятность опасного события, требуются другие меры и ответственные специалисты. Один порог нельзя бездумно переносить между задачами с разной ценой ошибки.

Для новичка полезная формулировка выглядит так: «На каких данных основана оценка? Что именно считается событием? Как проверялась точность?» Если ответ сводится к «мне кажется», процент не становится сильнее от двух знаков после запятой.

Разберите фразу «Я уверен на 97,3 процента, что в музее завтра свободный вход». Какие сведения лучше запросить вместо ещё одного процента? Прежде всего действующее объявление музея, дату и условия посещения. В этой задаче источник полезнее впечатления модели о собственной уверенности.

Двадцать прогнозов и осторожный вывод

В учебной серии есть двадцать прогнозов с оценкой вероятности 50 процентов. Событие произошло в одиннадцати случаях. Это близко к половине, но по такому небольшому набору нельзя объявить модель идеально откалиброванной. Если бы событие произошло восемь раз, одного этого тоже было бы мало для уверенного вывода о непригодности.

Посмотрите, одинаковы ли условия. Если часть прогнозов относится к одному сезону, а часть к другому, простое объединение может скрыть различия. Для первого упражнения достаточно понять: объём, состав и способ подсчёта влияют на оценку. Точная цифра без описания выборки остаётся неполным сообщением.

В отчёте полезно написать то, что действительно наблюдалось: «В двадцати выбранных случаях событие произошло одиннадцать раз; для оценки устойчивости результата нужны

дополнительные данные». Такая фраза не обещает больше, чем позволяет опыт, и оставляет понятный следующий шаг.

Глава 14 Контекст и память

Во время разговора человек помнит общую тему, предыдущие события и личный опыт. У цифрового помощника доступные сведения устроены иначе. Часть находится в текущей переписке, часть может храниться в специальных функциях памяти, часть извлекается из подключённых документов. Иногда ни один из этих механизмов не включён.

Контекстное окно — ограниченный объём информации, который модель может учитывать при обработке. В него могут входить пользовательский текст, инструкции системы, история и найденные материалы. Если переписка длинная, не всё прежнее обязательно остаётся доступным в исходном виде. Сервис может сокращать историю или выбирать отдельные фрагменты.

Отсюда знакомая ситуация: в начале диалога вы просили писать для восьмиклассника, а через некоторое время получили объяснение с незнакомыми терминами. Возможные причины различаются: потеря существенного условия, неоднозначный новый запрос или обычная ошибка следования инструкции. Практическое решение — кратко повторить актуальные требования.

Память сервиса не следует считать точной стенограммой всей жизни пользователя. Если система предлагает сохранять предпочтения, важно понять, что именно сохраняется и как этим управлять. Удаление одного разговора и удаление отдельной записи памяти могут быть разными действиями. Конкретные возможности определяются продуктом.

Для долгого учебного проекта полезна собственная карточка задачи. В ней достаточно указать тему, аудиторию, проверенные факты, принятые решения и оставшиеся вопросы. Перед новым сеансом передайте безопасную часть карточки помощнику. Это надёжнее, чем ожидать, что программа безошибочно восстановит все договорённости.

Например: «Мы готовим проект о расходе бумаги. Используем только вымышленные данные. Известно: за пять дней израсходовано 240 листов, учтено 6 кабинетов. Пока нельзя делать вывод о всей школе. Нужен план дальнейших наблюдений». Такой ввод короток, но содержит ограничения, без которых ответ легко станет преувеличением.

Длинный контекст сам по себе не гарантирует точности. В большом документе могут быть противоречия, устаревшие версии и посторонние указания. Пользователь должен назвать, какой файл основной и какие даты существенны. Если документов несколько, полезно попросить сначала перечислить расхождения, а уже затем делать выводы.

Задание: сократите описание школьной экскурсии до пяти предложений так, чтобы сохранились цель, дата, маршрут, неизвестные условия и запрет придумывать стоимость. Затем сравните с сокращением, в котором оставлены только красивые общие слова. Какое описание позволит другому человеку продолжить работу?

Глава 15 Поиск и генерация

Поисковая система находит материалы в своём индексе и показывает результаты, связанные с запросом. Генеративная модель создаёт ответ. Современный сервис может сочетать оба процесса, поэтому внешняя форма интерфейса не всегда подсказывает, откуда взялись сведения. Нужно различать функцию поиска и функцию написания текста.

Представим вопрос о режиме работы музея на ближайшие выходные. Здесь важны свежие сведения конкретного учреждения. Лучшее основание — действующее объявление на официальном сайте или подтверждение музея. Общее знание о типичном расписании не решает задачу. Даже верный ответ месяц назад может оказаться неверным сегодня.

Теперь другой вопрос: «Объясни, почему в русском языке бывают омонимы». Для первого объяснения важнее понятность и корректность понятий, чем обновление в реальном времени. Здесь языковая модель может быть полезным собеседником, а проверка выполняется по учебнику и словарю. Разный тип вопроса требует разного способа получения ответа.

Система с поиском иногда находит документ, извлекает фрагменты и использует их при генерации. Такой подход часто называют RAG — генерацией с дополнением найденными материалами. Он помогает связать ответ с конкретной базой знаний, однако ошибки остаются: найден не тот документ, пропущено условие или неправильно передана мысль источника.

Поэтому требуйте проверяемости. Полезно попросить указать название документа и место, на котором основан вывод. Если нужна дословная цитата, сравните её с оригиналом. Пересказ не должен выдаваться за цитирование. А одна найденная строка не обязательно отражает смысл всей страницы.

Поисковая выдача тоже не является удостоверением качества. Среди результатов могут быть реклама, пересказы, устаревшие страницы и материалы с заинтересованной подачей. Официальный источник не непогрешим во всех вопросах, но обычно лучше подходит для сведений о собственных правилах, реквизитах и услугах организации.

Простой порядок работы: определите, насколько информация меняется со временем; найдите подходящий источник; проверьте дату и предмет; только затем попросите помощника объяснить или сократить материал. Если порядок перевернуть, легко начать искать подтверждения уже понравившемуся ответу.

Практика: у вас есть старое объявление библиотеки и новое сообщение о ремонте. В старом указано обычное расписание, в новом — временное закрытие читального зала. Как должен выглядеть ответ посетителю? Почему простое объединение двух текстов без указания срока действия создаст путаницу?

Глава 16 Почему прочитанный файл ещё нужно проверить

Прикреплённый документ кажется моделью того, что человек видит на странице. Но программа может получить другой порядок текста, пропустить часть строк или неверно распознать символ. Кнопка успешной загрузки подтверждает передачу файла, а не полное и точное понимание его содержания.

В текстовом документе буквы хранятся как символы. В скане страницы они сначала представлены изображением. Чтобы получить текст, применяется оптическое распознавание, обычно обозначаемое OCR. Российский Vision OCR — пример такого инструмента. В задаче важен не бренд, а различие между чтением символов и их восстановлением по изображению.

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.