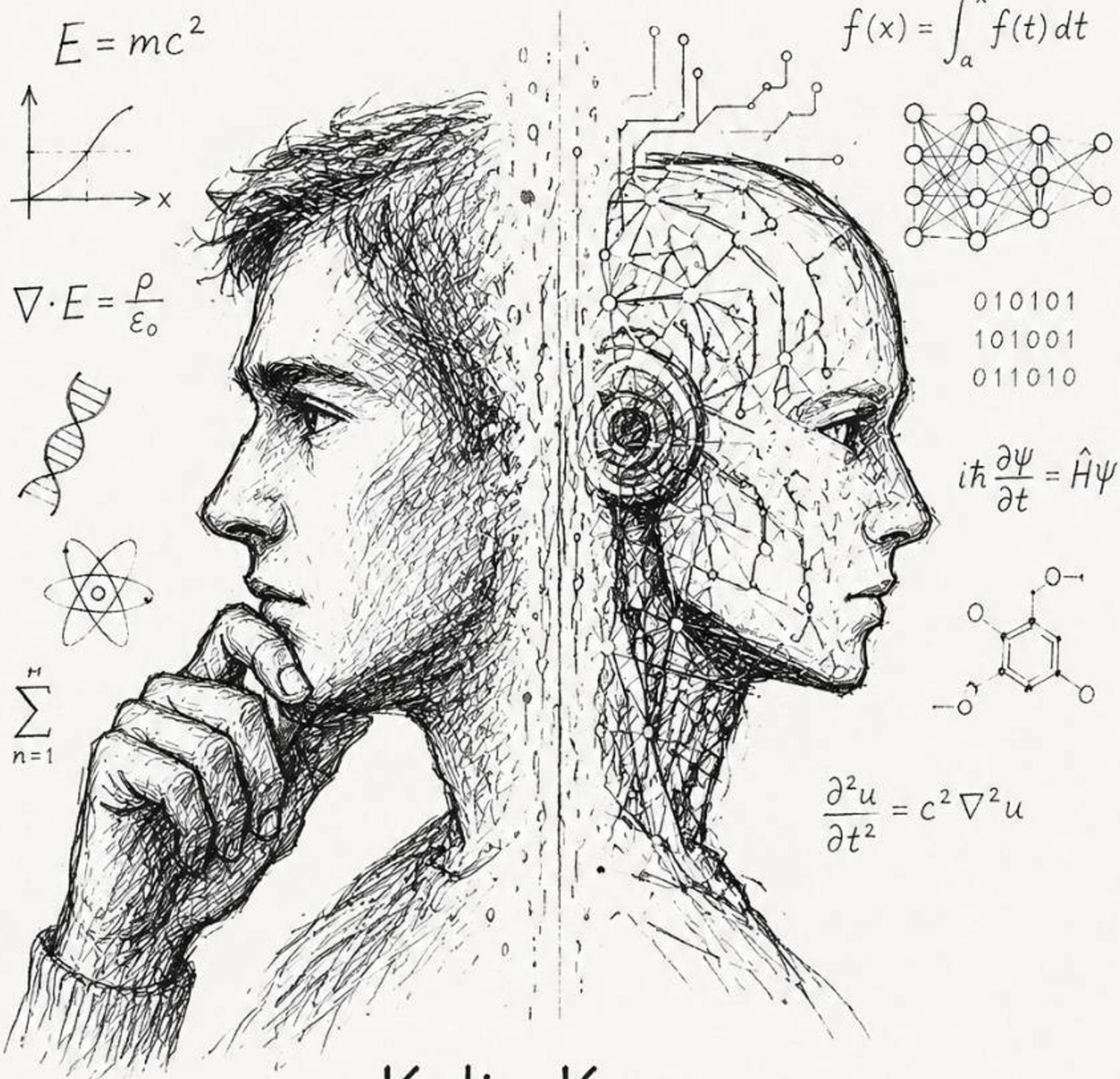
РАЗУМ, КОТОРЫЙ МЫ СОЗДАЛИ

Что произойдёт, если мы научим машины думать



Kalia Kavea

Kalia Kavea

Разум, который мы создали

«Автор»
2026

Kavea K.

Разум, который мы создали / K. Kavea — «Автор», 2026

Что произойдёт, если однажды машина начнёт находить то, чего человечество ещё не умеет искать? Искусственный интеллект создавался как инструмент — чтобы считать быстрее, работать с огромными объёмами информации и решать задачи, которые человеку становились всё сложнее. Но постепенно он начал делать нечто большее: находить закономерности, предлагать неожиданные решения, помогать учёным исследовать области, до которых человеческое мышление не всегда может добраться в одиночку. Эта книга о том, как человечество создало искусственный интеллект, зачем ему понадобился новый способ мышления и что происходит, когда машина становится не просто исполнителем, а участником поиска.

Содержание

Kalia Kavea
Разум, который мы создали

От автора

Я не начинала писать эту книгу с готовыми ответами.

На самом деле всё произошло наоборот.

Сначала был интерес. Потом — вопросы. А затем я начала читать: научные статьи, исследования, описания экспериментов, работы учёных, которые пытаются понять, на что уже способен искусственный интеллект и куда всё это может привести.

Чем больше я читала, тем меньше мне хотелось найти простой ответ.

Потому что за каждым новым достижением возникал следующий вопрос.

Если машина способна находить закономерности, которые человек не заметил, — что это меняет в науке?

Если она помогает создавать новые решения, — где заканчивается инструмент и начинается сотрудничество?

Если однажды мы научимся создавать системы, которые будут всё больше напоминать интеллектуальных существ, — что тогда вообще будет означать слово «разум»?

И постепенно я поняла, что не хочу писать книгу, которая заранее знает, чем всё закончится.

Мне хотелось сделать другое.

Собрать свои размышления.

Посмотреть на происходящее глазами человека, который живёт именно в этот момент — когда искусственный интеллект уже перестал быть фантастикой, но мы всё ещё не знаем, во что он в конечном счёте превратится.

Поэтому в этой книге рядом существуют наука и философия.

Здесь есть реальные исследования и эксперименты. Есть то, что уже удалось обнаружить и проверить. Есть результаты, которые можно воспроизвести и обсудить.

Но есть и вопросы, на которые у нас пока нет окончательных ответов.

И там, где я начинаю заглядывать в будущее, я не хочу выдавать предположение за факт.

Возможно, некоторые мои мысли окажутся ошибочными.

Возможно, будущее пойдёт совершенно другим путём.

Но разве это делает размышления о нём бессмысленными?

Мне кажется, нет.

Иногда нам нужно не знать ответ, а **осмелиться задать вопрос, на который ответа пока не существует.**

Эта книга появилась именно из такого любопытства.

Я читала о том, что уже умеет искусственный интеллект, и пыталась представить, что это может означать не только для машин, но и для нас.

Для науки.

Для человеческого мышления.

Для нашего представления о разуме.

И, возможно, для самого понятия знания.

Потому что, когда инструмент начинает помогать нам находить то, чего мы сами ещё не понимаем, меняется не только инструмент.

Меняется сам вопрос о том, **что значит знать.**

Я не знаю, станет ли когда-нибудь искусственный интеллект настоящим разумом.

И, возможно, сегодня мы ещё даже не умеем точно определить, что должно произойти, чтобы назвать систему разумной.

Но мне стало интересно другое:

а что, если мы сейчас находимся только в начале пути, смысл которого сами ещё не способны полностью увидеть?

Эта книга — моя попытка посмотреть на этот путь немного внимательнее.

Не предсказать его.

Не дать окончательный ответ.

А понять, **какие вопросы он уже заставляет нас задавать.**

Пролог

Мы создали нечто, чего ещё не существовало

В какой-то момент истории человечество сделало то, чего прежде не делало никогда.

Мы создали не просто новый инструмент.

Не новый двигатель. Не новый способ передвижения. Не машину, которая могла поднимать больше, двигаться быстрее или считать точнее человека.

Мы создали системы, способные работать с информацией и участвовать в процессах, которые прежде считались исключительно человеческими: распознавать закономерности, строить связи между разными фрагментами информации, предлагать решения и формулировать гипотезы.

Сначала это казалось продолжением уже знакомой истории.

Компьютеры давно превосходили человека в вычислениях. Они могли хранить огромные объёмы данных, выполнять сложнейшие операции и искать информацию за доли секунды.

Но постепенно искусственный интеллект начал делать нечто, что заставило нас посмотреть на него иначе.

Он стал находить закономерности.

Строить связи.

Предлагать решения.

Формулировать гипотезы.

Иногда — такие, которые человек не сформулировал заранее.

Именно здесь начинается история, которую мы пока не до конца понимаем.

На протяжении тысячелетий человеческое знание росло почти незаметно для самого человека.

Один человек мог знать растения своего леса. Другой — движение звёзд. Третий — свойства металлов. Четвёртый — устройство человеческого тела.

Потом появились города, письменность, библиотеки, университеты и научные общества.

Знаний становилось больше.

Затем — значительно больше.

А затем их стало столько, что ни один человек уже не мог охватить даже малую часть того, что знало человечество.

Это не было поражением человеческого разума.

Наоборот.

Это стало одним из величайших свидетельств его силы.

Мы научились разделять знания на области, создавать дисциплины и специализации, передавать результаты одного поколения следующему. Человек мог посвятить десятилетия одной области, одновременно оставаясь способным соединять её с другими областями знания. Учёные сотрудничивали, спорили, исправляли ошибки друг друга и постепенно собирали картину мира, которую невозможно было бы создать в одиночку.

Так возникла современная наука.

Но вместе с её успехом появилась странная проблема.

Чем больше мы узнавали, тем труднее становилось увидеть всё сразу.

Физик мог потратить годы на изучение одной закономерности. Биолог — десятилетия на исследование одной системы. Математик — годы на доказательство одной теоремы.

Каждый мог знать то, чего не знали другие.

Не потому, что другие уступали в интеллекте.

Просто человеческая жизнь имеет пределы.

Мы можем учиться десятилетиями. Можем стать специалистами сразу в нескольких областях. Можем объединять знания разных дисциплин и работать в командах.

Но всё человеческое знание всё равно остаётся распределённым между миллиардами отдельных умов.

И затем появилась машина, способная искать связи в огромных массивах информации совсем иначе.

Сначала мы создавали искусственный интеллект для вполне конкретных задач.

Распознавать изображения.

Переводить текст.

Играть в игры.

Предсказывать последовательности.

Находить закономерности в данных.

Помогать человеку работать быстрее.

Долгое время интеллект машины воспринимался как набор отдельных способностей.

Она могла хорошо считать, но не работала с языком.

Она могла распознавать изображения, но не могла вести содержательный анализ текста.

Она могла побеждать человека в игре, но это ещё не означало, что она понимала смысл самой игры.

Каждый новый успех казался ещё одной узкой победой технологии.

Затем границы начали размываться.

Машины стали всё лучше работать с языком.

А язык — это гораздо больше, чем средство общения.

Через него мы описываем физические законы, формулируем математические идеи, объясняем устройство клеток, записываем историю, создаём философские концепции, проектируем технологии и передаём опыт от одного человека к другому.

Язык связывает между собой огромные области человеческой деятельности.

И когда искусственный интеллект научился всё сложнее работать с этим пространством, произошло нечто важное.

Он получил возможность находить связи между огромным количеством информации, созданной людьми, — связи, которые не всегда были очевидны заранее.

И тогда возник вопрос, который мы, возможно, слишком долго откладывали:

Что произойдёт, если созданная нами система начнёт обнаруживать то, чего не заметили мы?

Мы привыкли считать открытие исключительно человеческим актом.

Учёный изучает данные.

Замечает странность.

Формулирует вопрос.

Предлагает гипотезу.

Проверяет её.

Иногда ошибается.

Иногда начинает сначала.

А иногда обнаруживает нечто, что меняет наше представление о мире.

Но что произойдёт, если часть этого процесса сможет выполнять не человек?

Если система обнаружит закономерность, которую исследователь не заметил?

Если предложит математическую конструкцию, которую никто не искал?

Если поможет найти молекулу с необычными свойствами?

Если соединит результаты нескольких научных областей и предложит гипотезу, которая прежде не возникала?

И если эксперимент подтвердит эту гипотезу — что именно произошло?

Мы получили новое знание.

Но кто был его источником?

Человек?

Машина?

Или мы впервые столкнулись с формой сотрудничества, для которой привычного представления об авторстве уже недостаточно?

Эта книга не пытается доказать, что искусственный интеллект станет доминирующей силой на Земле

И не пытается убедить вас в обратном.

Будущее ещё не написано.

Мы не знаем, каким окажется искусственный интеллект через десять, двадцать или пятьдесят лет. Не знаем, появится ли когда-нибудь сознание, которое можно будет назвать машинным. Не знаем, существуют ли принципиальные пределы развития таких систем. И не знаем даже, насколько хорошо понимаем собственный интеллект.

Но кое-что мы уже знаем.

Мы создали системы, способные участвовать в задачах, которые ещё недавно казались исключительно человеческими.

И, возможно, историческая важность этого шага заключается не в том, что машины стали похожи на людей.

Возможно, всё наоборот.

Впервые мы создали интеллект, который не обязан быть похожим на нас.

Он не обязан учиться так, как учимся мы.

Не обязан воспринимать мир так, как воспринимаем его мы.

Не обязан приходить к решениям тем же путём.

И именно поэтому сравнивать его с человеком становится одновременно необходимым и недостаточным.

Перед человечеством возникает необычная задача.

Нам предстоит понять не только то, **на что способен искусственный интеллект**.

Нам предстоит понять, **что происходит с нами, когда рядом появляется другой способ обрабатывать информацию, искать закономерности и приходить к решениям.**

Потому что, возможно, величайшее открытие эпохи искусственного интеллекта окажется не открытием о машинах.

А открытием о самих себе.

О том, что такое интеллект.

Что значит понимать.

Что значит создавать новое.

Что именно происходит в тот момент, когда найденное решение существует, но путь к нему не похож на человеческое рассуждение.

И, наконец, что значит быть человеком — в мире, где способность решать всё более сложные интеллектуальные задачи больше не является нашей исключительной территорией.

Возможно, прежде чем спросить, каким станет искусственный интеллект, нам придётся заново спросить себя:

что именно мы называем интеллектом?

Глава 1

Мы всегда пытались превзойти собственные пределы

Представьте человека, который впервые взял в руки камень и понял, что теперь может сделать то, чего раньше не мог.

Расколоть орех.

Разбить кость.

Обработать другой камень.

Защититься от хищника.

Сам камень не стал частью его тела. Но человеческое действие изменилось.

В этом простом моменте скрывается одна из самых важных особенностей нашей истории.

Мы никогда не ограничивались тем, чем были наделены природой.

Мы создавали то, чего у нас не было.

Инструмент становился продолжением руки.

Колесо — продолжением движения.

Лодка — продолжением способности пересекать воду.

И каждый раз граница между тем, что человек мог сделать сам, и тем, что он мог сделать с помощью созданного им инструмента, отодвигалась немного дальше.

Но однажды человечество начало расширять не только тело.

Мы начали расширять память.

Первым великим интеллектуальным инструментом была не вычислительная машина.

Это была письменность.

Пока знания существовали главным образом в памяти людей, их сохранение зависело от способности одного поколения передать опыт следующему. Человек мог рассказать другому, где найти лекарственное растение, как обработать металл или как ориентироваться по звёздам.

Но память несовершенна.

Детали забываются.

Истории меняются.

Ошибки передаются вместе с фактами.

Письменность позволила сделать нечто принципиально новое: **оставить информацию вне человеческой памяти.**

Можно было записать наблюдение и передать его человеку, которого никогда не встретишь.

Сохранить закон.

Описание растения.

Математическую задачу.

Историю.

Карту.

Наблюдение за небесным телом.

Знание больше не должно было исчезать вместе с человеком, который им обладал.

Это изменило саму природу накопления знаний.

Теперь новое поколение могло начинать не с того места, где начинало предыдущее, а с того, где оно остановилось.

Но чем больше знаний удавалось сохранить, тем сложнее становилось ими управлять.

Появились архивы.

Библиотеки.

Каталоги.

Системы классификации.

Человечество создавало всё новые способы не только сохранять информацию, но и находить её.

Затем появился печатный станок.

В середине XV века технология Иоганна Гутенберга резко увеличила скорость воспроизведения текстов в Европе. Книги перестали быть исключительно результатом кропотливого ручного переписывания и начали распространяться в значительно больших масштабах.

Но значение этого изменения было гораздо глубже самой технологии печати.

Представьте учёного, который делает наблюдение.

В мире рукописей его идея может остаться известной небольшому кругу людей.

В мире массового книгопечатания она получает возможность путешествовать гораздо дальше своего автора.

Один человек пишет книгу.

Её читают тысячи.

Один исследователь описывает эксперимент.

Другие могут ознакомиться с его результатами и попытаться повторить их.

Математик публикует доказательство.

Следующий учёный начинает работу уже с этого результата.

Знание становится накопительным.

Каждое поколение получает возможность строить новое поверх уже созданного.

Именно так постепенно возникла одна из величайших интеллектуальных конструкций человечества — современная наука.

Но её успех породил новую проблему.

Чем больше становилось знаний, тем труднее было одному человеку охватить их все.

Началась специализация.

Физик мог посвятить годы изучению свойств материи.

Биолог — живых систем.

Химик — взаимодействий между веществами.

Математик — абстрактных структур.

Инженер — превращению научных идей в работающие технологии.

Это не означало, что человек больше не способен понимать разные области.

Напротив, многие выдающиеся исследователи работали на пересечении дисциплин.

Но существовала другая граница.

Масштаб человеческого знания рос быстрее, чем способность отдельного человека изучить его целиком.

Можно было стать специалистом в нескольких областях.

Можно было всю жизнь расширять свои знания.

Можно было сотрудничать с другими исследователями.

Но невозможно было лично прочитать и глубоко осмыслить всё, что создавала современная цивилизация.

И это было не поражением человека.

Это было следствием его успеха.

Мы создали столько знаний, что теперь сам их масштаб требовал новых инструментов.

Следующий такой инструмент оказался уже не бумажным.

Он был электронным.

Одной из первых больших задач вычислительных машин стало то, с чем человек всегда справлялся хуже всего: выполнение огромного количества точных операций.

Человек способен распознавать лица, понимать контекст, строить планы и замечать неожиданные события. Но он не предназначен для выполнения миллионов одинаковых операций с абсолютной точностью и постоянной скоростью.

Машина — предназначена.

Сначала компьютеры выполняли расчёты, которые человеку потребовали бы огромного количества времени.

Затем они начали работать с задачами, которые становились практически невозможными для ручного выполнения.

Но компьютер оставался инструментом.

Человек определял процедуру.

Машина выполняла её.

Именно здесь проходит важная граница между традиционным программированием и тем, что позже станет машинным обучением.

В обычной программе человек заранее описывает правила.

Если происходит одно событие — выполнить определённое действие.

Если число превышает заданное значение — применить определённую операцию.

Если пользователь нажимает кнопку — запустить процедуру.

Машина следует инструкции.

Но что делать, если мы сами не знаем всех правил?

Представьте задачу распознавания кошки на фотографии.

Мы могли бы попытаться описать её математически.

Четыре лапы.

Два глаза.

Шерсть.

Уши определённой формы.

Хвост.

Но достаточно показать системе сфинкса, котёнка, животное в необычной позе или фотографию, сделанную в темноте, — и список заранее написанных правил быстро становится недостаточным.

Проблема не в том, что мы не способны описать кошку.

Проблема в том, что реальный мир слишком разнообразен, чтобы полностью свести его к заранее составленному перечню инструкций.

И тогда возникает другой подход.

Не объяснять машине каждое правило.

Дать ей множество примеров и позволить алгоритму находить закономерности в данных.

Так постепенно сформировалась идея машинного обучения.

Она изменила саму роль программиста.

Вместо того чтобы вручную описывать каждую возможную ситуацию, человек мог создавать алгоритм обучения, выбирать данные и задавать критерий, по которому система должна улучшать результат.

Машине больше не обязательно было сообщать каждый шаг.

Ей можно было показать множество примеров и позволить вычислительной системе самостоятельно настроить внутренние параметры так, чтобы лучше выполнять поставленную задачу.

Это был фундаментальный сдвиг.

Но он не произошёл мгновенно.

История искусственного интеллекта началась задолго до современных нейронных сетей.

В 1950 году Алан Тьюринг опубликовал статью *Computing Machinery and Intelligence*, в которой исследовал вопрос о возможности машинного мышления и предложил заменить расплывчатый вопрос «могут ли машины мыслить?» более конкретным экспериментальным подходом.

Через несколько лет появилась идея, которая дала новой области исследований её название.

В 1955 году Джон Маккарти, Марвин Минский, Натан Рочестер и Клод Шеннон подготовили предложение о летнем исследовательском проекте в Дартмутском колледже. В нём использовалось выражение **artificial intelligence** — «искусственный интеллект».

Летом 1956 года участники проекта собрались для обсуждения возможности исследовать обучение, рассуждение, решение задач и другие процессы, которые обычно связывали с человеческим интеллектом.

Их предположение было смелым.

Они считали, что различные аспекты обучения и интеллекта в принципе могут быть описаны настолько точно, чтобы машина могла их воспроизводить или моделировать.

Это не было доказательством.

Не было готовой технологией.

И уж тем более не было гарантией быстрого создания искусственного разума.

Это была исследовательская гипотеза.

Но именно такие гипотезы иногда становятся началом длинных научных историй.

Первые десятилетия развития ИИ были полны противоречий.

Одни исследователи строили системы на логике и символах.

Другие пытались создавать модели, вдохновлённые устройством нервной системы.

Появлялись программы, способные решать математические задачи, играть в игры и работать с языком.

Некоторые результаты выглядели впечатляюще.

Некоторые обещания оказались преждевременными.

Технологии часто не соответствовали ожиданиям.

Наступали периоды разочарования, когда финансирование сокращалось, интерес к отдельным направлениям падал, а обещанный прогресс не наступал.

Но научные идеи редко исчезают только потому, что их первое воплощение оказалось неудачным.

Иногда проблема находится не в самой идее, а в недостатке вычислительной мощности.

Иногда не хватает данных.

Иногда ещё не существует подходящего алгоритма.

А иногда исследователи просто задают природе вопрос, на который технология пока не умеет отвечать.

Тем временем менялся сам мир.

Компьютеры становились быстрее.

Память — дешевле.

Данные — доступнее.

А компьютеры начали соединяться друг с другом.

Появление интернета изменило масштаб человеческой информационной среды.

Раньше компьютер мог быть мощным инструментом отдельного человека или организации.

Теперь миллиарды устройств получили возможность обмениваться информацией.

Человечество словно построило внешнюю систему памяти и связало её миллиардами каналов.

Мы начали создавать и передавать информацию с такой скоростью, которая ещё недавно казалась невозможной.

Каждый текст, фотография, научная статья, измерение, программа и цифровая транзакция могли становиться частью огромного массива данных.

И здесь возник новый парадокс.

Мы научились создавать информацию быстрее, чем человек способен её прочитать.

Научная литература продолжала расти.

Цифровые архивы увеличивались.

Количество данных, которые оставляют после себя люди и созданные ими системы, стало огромным.

Но способность отдельного человека читать, сравнивать и осмысливать всё это не увеличивалась с той же скоростью.

Сутки по-прежнему состоят из двадцати четырёх часов.

Человеческая жизнь по-прежнему конечна.

И это означало, что следующий шаг должен был быть связан уже не только с хранением информации.

Нужна была система, способная **искать структуру внутри её масштаба.**

К этому моменту несколько линий развития встретились.

У человечества были огромные массивы данных.

Были компьютеры, способные выполнять колоссальное количество операций.

Были десятилетия исследований в области алгоритмов и статистики.

И существовала идея, которая постепенно становилась всё более важной:

машину необязательно программировать для каждого отдельного случая.

Её можно обучать на примерах.

Именно сочетание этих факторов позволило машинному обучению перейти из относительно узкой исследовательской области в одну из центральных технологий современного ИИ.

Нейронные сети становились глубже и эффективнее.

Методы обучения улучшались.

Вычислительные мощности росли.

Объёмы доступных данных увеличивались.

И системы начали справляться с задачами, которые ещё недавно казались слишком сложными для машин.

Распознавать речь.

Анализировать изображения.

Переводить языки.

Играть в сложные игры.

Генерировать текст.

Работать с программным кодом.

А затем — соединять несколько таких способностей в одной системе.

Это был уже другой тип инструмента.

Не потому, что машина внезапно стала человеком.

А потому, что человек всё меньше задавал ей каждый конкретный шаг.

Мы могли поставить задачу.

Предоставить данные.

Определить критерий результата.

А часть пути между началом и ответом система находила сама.

И именно здесь возникает вопрос, который определит следующую главу нашей истории.

Что на самом деле происходит, когда машина учится?

Если человек не объясняет ей каждое правило, то что именно она извлекает из примеров?

Как информация превращается в внутренние представления?

Почему система иногда находит закономерности, которые мы не задавали ей напрямую?

И где проходит граница между тем, что мы называем вычислением, и тем, что начинаем называть обучением?

Чтобы ответить на эти вопросы, нам теперь придётся сделать то, чего мы раньше не делали.

Не просто посмотреть на то, **что машина делает**.

А заглянуть внутрь самого процесса, благодаря которому она этому научилась.

Глава 2

Как научить машину учиться

Представим простую задачу.

Перед нами тысячи фотографий.

На одних — кошки.

На других — собаки.

Мы хотим, чтобы компьютер научился различать их.

На первый взгляд задача кажется элементарной.

Человек обычно узнаёт кошку почти мгновенно. Мы обращаем внимание на форму головы, глаза, уши, шерсть, тело, движения — иногда даже тогда, когда видим животное всего несколько секунд.

Но если попытаться объяснить компьютеру, **что именно делает кошку кошкой**, задача неожиданно становится сложной.

Какие должны быть глаза?

Какой длины уши?

Обязательно ли наличие усов?

Что делать с бесшёрстной кошкой?

А если животное сидит спиной?

А если фотография размыта?

А если кошка лежит в траве и видна только половина её тела?

Можно написать сотни правил.

Но рано или поздно появится фотография, для которой этих правил окажется недостаточно.

И тогда возникает другая идея.

Вместо того чтобы объяснять машине каждое правило, можно показать ей множество примеров.

Представим, что мы даём системе десять тысяч изображений.

На каждом изображении заранее указано, что находится перед ней:

кошка.

собака.

Система начинает сравнивать изображения.

Она ещё не знает, какие признаки важны.

Не знает, что такое уши.

Не знает, что такое морда.

Не знает, что перед ней вообще животное.

Она получает данные и пытается найти в них закономерности, которые помогут отличать один класс изображений от другого.

Сначала результаты будут плохими.

Система может ошибаться снова и снова.

Но после каждой ошибки можно изменить её внутренние параметры так, чтобы следующий результат оказался немного лучше.

Именно здесь появляется фундаментальная идея машинного обучения:

система не получает готовое описание решения — она изменяет себя в процессе обучения так, чтобы лучше соответствовать данным и поставленной задаче.

Это принципиально отличается от обычной программы.

В традиционном программировании человек формулирует правила, а компьютер выполняет их.

В машинном обучении человек задаёт структуру процесса обучения, предоставляет данные и определяет, что считать хорошим результатом.

Но конкретные значения огромного количества внутренних параметров система настраивает сама.

Можно представить это очень упрощённо.

У системы есть огромное количество внутренних чисел — параметров.

В начале их значения подобраны так, что система почти ничего не умеет.

Она получает изображение.

Делает предсказание.

Сравнивает его с правильным ответом.

Возникает ошибка.

Затем алгоритм вычисляет, как изменить параметры, чтобы уменьшить эту ошибку.

После этого система получает следующий пример.

И снова делает предсказание.

Снова ошибается.

Снова корректирует параметры.

Так происходит не десятки и не сотни раз.

В современных системах обучение может включать огромное количество примеров и вычислений.

Постепенно параметры изменяются.

Система начинает находить устойчивые закономерности.

И если обучение прошло успешно, она способна работать не только с теми изображениями, которые уже видела, но и с новыми.

Это называется **обобщением**.

Именно здесь происходит одна из самых важных вещей.

Система не просто запоминает ответы.

Она должна научиться использовать найденные закономерности на новых данных.

Но что именно она находит?

Это уже гораздо интереснее.

Представьте, что мы смотрим на фотографию кошки.

Человек может сказать:

«Я вижу глаза, уши, нос, шерсть и форму тела».

Но нейронная сеть не обязана представлять изображение таким образом.

На первых уровнях обработки она может реагировать на простые структуры изображения — например, на границы, контраст и определённые комбинации пикселей.

На следующих уровнях могут формироваться более сложные представления.

Линии объединяются в формы.

Формы — в более сложные структуры.

А сложные структуры могут становиться частью представления целого объекта.

Это очень упрощённая картина того, что происходит в глубокой нейронной сети.

Но она помогает понять главное:

система не получает от человека готовую карту того, какие признаки искать.

Эта карта формируется в процессе обучения.

Именно поэтому нейронные сети стали настолько важны.

Их устройство частично вдохновлено тем, как мы описываем работу биологических нейронных систем, хотя современные искусственные нейронные сети значительно отличаются от человеческого мозга.

Вместо одного огромного вычислительного механизма они состоят из множества связанных элементов, организованных в слои.

Каждый элемент выполняет относительно простое математическое преобразование.

Но когда таких элементов становится очень много и они соединены определённым образом, система способна моделировать чрезвычайно сложные зависимости.

Название «нейронная сеть» звучит почти биологически.

Но важно помнить: это математическая модель.

Она не является цифровой копией человеческого мозга.

И она не обязана работать так же, как человеческий разум.

Это самостоятельный способ построения вычислительной системы.

С изображениями эта идея уже была мощной.

Но затем перед исследователями возникла другая задача.

Язык.

На первый взгляд текст должен быть проще фотографии.

Он уже состоит из отдельных символов и слов.

Но смысл языка оказывается намного сложнее.

Возьмём слово «ключ».

Это может быть ключ от двери.

Ключ к задаче.

Ключ как источник воды.

Или ключ музыкальный.

Само слово остаётся одинаковым.

Меняется контекст.

Человек почти автоматически учитывает слова вокруг него и понимает, какое значение здесь подходит.

Для машины это необходимо было превратить в вычислительную задачу.

Система должна была учитывать не только отдельные слова, но и отношения между ними.

И чем длиннее становился текст, тем важнее становилось понимать, **какая часть контекста связана с какой другой частью.**

Здесь произошёл один из наиболее важных сдвигов в развитии современных языковых моделей.

В 2017 году исследователи представили архитектуру **Transformer**.

Её ключевой механизм — attention, или механизм внимания, — позволял модели эффективнее учитывать связи между различными элементами последовательности.

Если очень упростить, модель получила способ оценивать, **на какие другие части входного текста стоит обратить больше внимания при обработке конкретного элемента.**

Это особенно важно для языка.

Потому что смысл одного слова может зависеть от слов, находящихся далеко от него.

Предложение может начинаться с одной идеи, продолжаться несколькими уточнениями, а завершаться словом, смысл которого становится понятным только после всего предыдущего контекста.

Transformer позволил обрабатывать такие зависимости значительно эффективнее, чем многие предыдущие подходы.

И эта архитектура стала одним из фундаментальных компонентов современного развития языковых моделей.

Но архитектура сама по себе не создаёт интеллект.

Нужны данные.

Нужны вычислительные ресурсы.

Нужен процесс обучения.

И здесь возникает масштаб, который трудно представить интуитивно.

Современная модель может содержать огромное количество параметров.

Во время обучения эти параметры многократно изменяются.

Система получает последовательности данных, делает предсказания, вычисляет ошибку и корректирует параметры.

Для языковой модели одной из фундаментальных задач может быть предсказание следующего элемента последовательности.

Например:

«Солнце взошло над…»

Система должна оценить, какое продолжение наиболее вероятно.

На первый взгляд это может показаться слишком простым.

Но если повторить подобный процесс на огромном количестве текстов, задача становится значительно сложнее.

Чтобы хорошо предсказывать следующий фрагмент текста, модели приходится учитывать грамматику, структуру предложений, устойчивые выражения, отношения между словами и множество других закономерностей языка.

А при достаточно большом масштабе в процессе обучения начинают формироваться представления, позволяющие выполнять гораздо более широкий круг задач.

И здесь появляется важное различие между **запоминанием** и **обобщением**.

Если система просто запоминает огромное количество примеров, она может прекрасно работать с тем, что уже встречала.

Но настоящая ценность обучения проявляется тогда, когда система сталкивается с новым примером и всё же способна применить ранее найденные закономерности.

Именно поэтому исследователи так внимательно изучают способность моделей обобщать.

Модель не должна просто помнить.

Она должна использовать приобретённые представления в новых ситуациях.

Но здесь возникает и проблема.

Система может найти закономерность, которая хорошо работает на обучающих данных, но плохо работает за их пределами.

Она может уловить случайную зависимость.

Может переобучиться.

Может дать уверенный ответ там, где данных недостаточно.

Поэтому способность модели выдавать убедительный результат ещё не означает, что этот результат правильный.

Это одна из причин, по которой современные системы необходимо проверять.

И всё же масштаб меняет характер происходящего.

Когда модель имеет огромное количество параметров, обучается на огромных массивах данных и получает достаточно вычислительных ресурсов, она может демонстрировать способности, которые не были вручную прописаны разработчиком в виде отдельных правил.

Это не означает, что программист создал внутри неё маленького человека.

Не означает, что система обладает сознанием.

И не означает автоматически, что она понимает мир так же, как понимаем его мы.

Но означает кое-что важное.

Человек перестал вручную описывать каждую закономерность, которую должна использовать машина.

Мы создаём процесс.

Предоставляем данные.

Определяем цель.

Настраиваем архитектуру.

Запускаем обучение.

А затем наблюдаем за тем, какие внутренние представления и способности возникают в результате.

Именно здесь современные системы становятся особенно интересными.

Потому что мы уже не всегда можем объяснить их поведение простым списком правил.

Представим теперь, что мы обучили систему не различать кошек и собак, а работать с научными данными.

Она получает результаты экспериментов.

Научные статьи.

Измерения.

Математические зависимости.

Описание известных веществ.

Результаты предыдущих исследований.

Мы можем поставить перед ней конкретную задачу.

Но мы не обязательно заранее знаем, какие связи она обнаружит внутри этих данных.

И тогда возникает новая возможность.

Машина может стать не только исполнителем заранее известных инструкций.

Она может стать инструментом **поиска в пространстве возможностей**, которое слишком велико для ручного перебора.

Но здесь начинается совершенно другая проблема.

Если система предлагает результат, которого мы не ожидали, откуда мы знаем, что она действительно обнаружила что-то важное?

Может быть, она нашла настоящую закономерность.

Может быть, случайную зависимость.

Может быть, использовала скрытый признак, которого исследователь не заметил.

А может быть, просто ошиблась.

Поэтому способность генерировать результат — это только начало.

Следующий вопрос гораздо сложнее:

может ли машина найти не просто ответ, а что-то новое, чего мы сами не искали?

Именно здесь заканчивается история о том, **как машина учится**

Конец ознакомительного фрагмента.