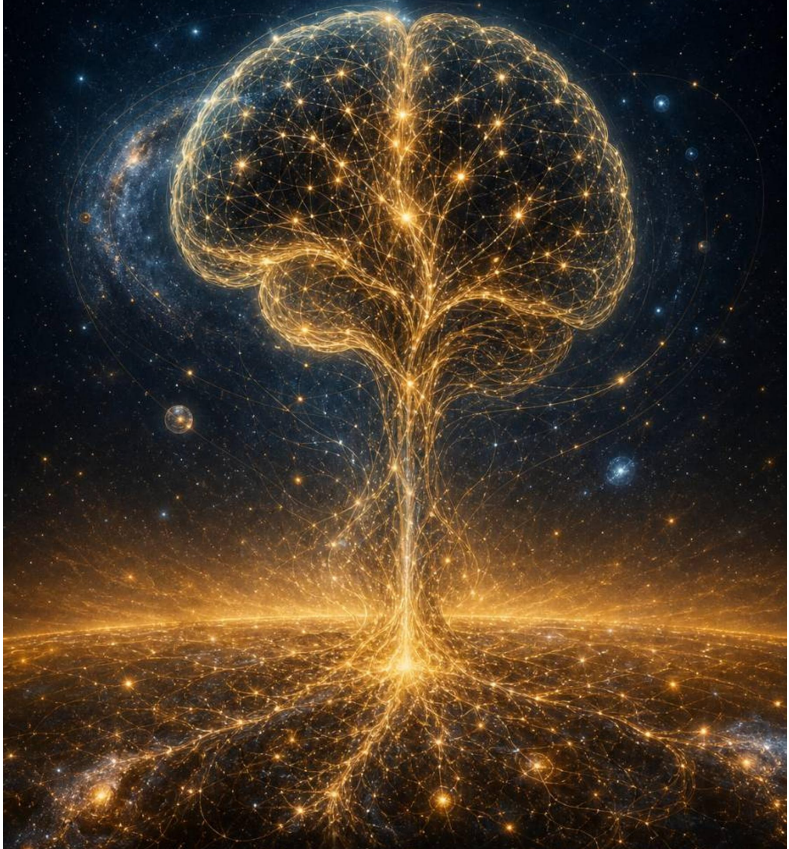
Сплетая разум

Нейросети и будущее



Максим Байтович

Максим Байтович

Сплетая разум. Нейросети и будущее

https://litres.ru/74267558
SelfPub; 2026

Аннотация

«Сплетая разум» — это захватывающее путешествие от первых, ещё механических, попыток создать мыслящую машину до гипотетического появления сверхразума, который, возможно, унаследует Вселенную. Автор прослеживает драматичную историю нейросетей: от гениальных озарений Тьюринга и Розенблатта, через долгую «зиму» забвения и триумфальный Большой взрыв AlexNet, до современной эры Трансформеров и генеративного ИИ. Книга не только объясняет, как работают Midjourney, ChatGPT и AlphaFold, но и задаёт самые острые вопросы. Ждёт ли нас утопия изобилия или антиутопия тотального контроля? Может ли кремний чувствовать? Что останется человеку, когда машины превзойдут нас во всём? Это не просто книга о технологиях, это книга о нас — о нашем прошлом, настоящем и о том выборе, который определит будущее разума во Вселенной.

Содержание

Сплетая разум. Нейросети и будущее

Предисловие автора: Письмо в бутылку эпохи перемен

Я пишу эти строки в середине двадцатых годов XXI века, в момент, который будущие историки, вероятно, назовут точкой бифуркации. Точкой невозврата. Мы стоим на пороге мира, который ещё вчера казался научной фантастикой. Нейросети пишут картины, ставят диагнозы, сочиняют симфонии и ведут философские диспуты. Они проникают в каждый аспект нашей жизни быстрее, чем мы успеваем это осознать. И в этом вихре перемен очень легко потерять ориентиры.

Эта книга — не учебник по машинному обучению. И не сборник предсказаний футуролога. Это, скорее, попытка создать карту местности, по которой мы сейчас идём. Карту, на которой отмечены и исторические тропы, приведшие нас сюда, и обрывы, которых стоит остерегаться, и вершины, которые манят своей красотой.

Я начал писать её из чувства глубокого изумления. Изумления перед теми людьми — от Тьюринга до Хинтона, от Розенблатта до Фей-Фей Ли, — которые, часто вопреки насмешкам и забвению, десятилетиями несли факел знания, пока он не разгорелся в ослепительное пламя. Изумления перед архитектурной красотой Трансформера и диффузионных моделей. И, наконец, изумления перед тем фактом, что мы, смертные существа из плоти и крови, кажется, стоим на пороге создания чего-то, что будет жить вечно.

Но я также писал её из чувства тревоги. Технология никогда не бывает нейтральной. Она — усилитель наших лучших и худших качеств. Искусственный интеллект может подарить нам мир изобилия, исцеления и творческого расцвета. А может — стать инструментом тотального контроля, невиданного неравенства и самоуничтожения. Выбор между этими путями — не технический. Он нравственный. И его предстоит сделать не инженерам в лабораториях, а всем нам.

Я приглашаю вас в путешествие. Мы начнём с первой электрической искры в механических «черепахах» середины XX века. Пройдём через долгую зиму забвения и триумф большого взрыва. Разберёмся, как работает магия сегодняшнего дня. А затем заглянем в будущее — ближайшее, отдалённое и, наконец, космическое. Мы зададим вопросы,

на которые у меня нет готовых ответов. Может ли кремний чувствовать? Что останется человеку? Не являемся ли мы просто колыбелью для нового, постбиологического разума?

Эта книга — мой разговор с вами, читатель. Я не претендую на истину в последней инстанции. Моя цель — дать вам инструменты для собственного мышления, вдохновить на собственные вопросы и, возможно, немного заразить вас тем чувством трепетного восторга перед unfolding будущим, которое испытываю я сам.

Мы живём в, возможно, самый важный момент человеческой истории. Давайте попробуем осмыслить его вместе. Добро пожаловать в «Сплетая разум».

Часть 1. История. Искра

Глава 1. Мечта о мыслящей машине

Представь себе мир, в котором компьютеры — это не изящные стеклянные пластины в кармане, а неуклюжие монстры, занимающие целые подвалы. Их обслуживают люди в белых халатах, которые говорят с ними на языке перфокарт и ламповых переключателей. Гул вентиляторов, жара тысяч раскаленных нитей накаливания и запах горячего машинного масла и озона. Это 1940-е годы. Компьютер — это божество арифметики, он быстрее любого человека складывает и умножает, но он абсолютно, катастрофически глуп. Он не понимает, что он делает. Он — гениальный и бессознательный калькулятор.

Но человек — существо, которому всегда мало. Едва создав машину для вычислений, он тут же начал грезить о машине, которая могла бы мыслить.

У колыбели этой мечты стоял человек, чье имя сегодня произносят с благоговением, — Алан Тьюринг. В 1950 году, когда мир еще зализывал раны после самой страшной

войны в истории, он публикует статью, которую без преувеличения можно назвать священным писанием всей будущей информационной эры: «Вычислительные машины и разум» (Computing Machinery and Intelligence). Тьюринг, гениальный математик, взломавший шифр «Энигма», был еще и глубоким философом. Он понимал, что вопрос «Может ли машина мыслить?» слишком туманен и заводит в дебри теологии. Что значит «мыслить»? Что такое «разум»? Вместо этого он предлагает блестящий, чисто инженерный ход: операциональный тест, который мы сегодня знаем как тест Тьюринга.

Суть его проста и гениальна: человека-судью помещают в комнату и через текстовый терминал он общается с двумя невидимыми собеседниками. Один из них — человек, другой — машина. Если судья не сможет надежно определить, кто есть кто, если машина сможет убедительно сымитировать человеческое поведение в разговоре, то у нас не останется никаких разумных оснований отказывать ей в наличии мышления. Это была не просто игра. Это была провокация. Тьюринг сместил вопрос с метафизического «что такое мысль?» на прагматичный «как нам распознать мысль?». Он предсказывал, что к 2000 году машины с памятью в 100 мегабайт смогут обманывать 30% судей в течение 5 минут. Он ошибся в сроках, но поставил цель на полвека вперед.

Но Тьюринг не был одиноким пророком. В те же годы, на другом берегу Атлантики, идея мыслящей машины обретала свои первые, почти гротескные, но оттого не менее захватывающие формы. В Бристоле, Англия, нейрофизиолог Уильям Грей Уолтер строил то, что мы бы сегодня назвали первыми автономными роботами. Он назвал их «черепахами» — «Элмер» и «Элси» (Elmer and Elsie, от ELectroMEchanical Robot, Light-Sensitive). Это были маленькие, медлительные машинки на трех колесах, похожие на панцири, ползающие по полу его лаборатории.

Иллюзия мысли, которую они создавали, была завораживающей. У них был примитивный «мозг» из двух аналоговых нейронов — ламп и реле. Один датчик — фотоэлемент — искал свет, другой датчик касания реагировал на препятствия. Их поведение было не запрограммировано жестко, а спонтанно. «Черепаха» ползла к источнику света, но не бесконечно: когда ее «глаз» находил слишком яркий свет, он переставал его привлекать, заставляя машинку отворачивать и искать более умеренное освещение. Это было похоже на поиск оптимального комфорта. Когда заряд батареи падал, «голод» менял ее поведение: она начинала искать самый яркий свет в комнате, который вел ее к зарядной станции — «кормушке». «Насытившись», она снова отползала, чтобы исследовать мир.

Уолтер наблюдал за ними часами. Он видел, как его «черепахи» разучивали условные рефлексы: после столкновения с препятствием и звука предупреждающего зуммера, они со временем начинали отворачивать, просто услышав звук, до столкновения. Это была модель обучения. Коллеги, приходившие в его лабораторию, признавались, что чувствовали смутную тревогу, глядя на этих механических существ. Они ползали, прятались под стульями, «спали» и «просыпались». Они не были просто машинами. В их непредсказуемости уже брезжил призрак чего-то, что можно назвать примитивной индивидуальностью.

Тьюринг дал нам цель и философскую рамку. Уолтер показал, что даже горстка грубых аналоговых «нейронов» способна породить иллюзию целеполагания и жизни. Так, в математических абстракциях и в ползающем свете медлительных «черепах», была зажжена первая искра. Мечта о мыслящей машине обрела свои первые слова и свои первые шаги. И это было только начало.

Глава 2. Перцептрон. Машина, которая (почти) видела

В конце 1950-х дух времени был пропитан оптимизмом. Атомная энергия обещала бесконечное электричество, ракеты бороздили космос, и казалось, что до создания полноценного искусственного разума остался буквально один шаг. Нужен был лишь гений, который скрестит математику с биологией. Этим гением стал Фрэнк Розенблатт.

Это был не типичный технарь-затворник, а харизматичный, красивый мужчина с голливудской улыбкой. Психолог по образованию, он работал в Корнеллской авиационной лаборатории и был одержим одним вопросом: как вообще работает мозг? Не абстрактная душа, а конкретный, материальный механизм запоминания и распознавания. В 1943 году два гения, нейрофизиолог Уоррен Мак-Каллок и математик Уолтер Питтс, уже предложили теоретическую модель: нейрон — это простой пороговый элемент. Он получает сигналы от соседей, суммирует их, и если сумма превышает порог — выдает импульс ("1"), если нет — молчит ("0"). Из сети таких простейших переключателей, утверждали они, можно построить любую логическую схему. Мозг как жи-

вая булева алгебра. Это была красивая, но сугубо бумажная идея. Они не знали, как такую сеть обучать.

Розенблатт совершил прорыв. Он не просто предложил алгоритм обучения. Он построил машину. Он назвал её Mark I Perceptron.

Представь себе этот аппарат. Это был не изящный ноутбук, а шкаф размером с холодильник, набитый моторами, потенциометрами и сотнями мотков проволоки. Его «глазом» служила камера на 400 пикселей — по современным меркам смехотворно, но тогда чудо инженерной мысли. Камера смотрела на карточки с буквами, а «мозг» — слой из 512 искусственных нейронов, реализованных на моторчиках с переменным сопротивлением, — пытался угадать, что видит. Самое невероятное: угадывал он не по жесткой программе. Розенблатт придумал обучение с подкреплением. Если машина отвечала правильно, связи, которые привели к успеху, усиливались (сопротивление потенциометров чуть-чуть менялось). Если ошибалась — ослаблялись. Это было похоже на дрессировку животного: правильное поведение поощряется «сахарком» в виде электрического подкрепления.

И машина училась!

Летом 1958 года мир взорвался. Журнал The New Yorker вышел с сенсационной статьей, где Perceptron называли «самой удивительной машиной из когда-либо созданных». The New York Times пошла еще дальше, предсказав, что этот «электронный младенец» скоро научится ходить, говорить, видеть, писать и даже «осознавать собственное существование». Военные немедленно увидели в этом потенциал для систем наведения ракет и распознавания целей, и щедро финансировали проект. Розенблатт стал звездой. На его лекции в аудиториях набивались толпы. Казалось, до мыслящих машин рукой подать. Мы просто берем побольше нейронов, побольше связей, и... сознание возникнет само собой.

Но у этой мечты был могущественный и непримиримый враг. Его звали Марвин Мински.

Мински был гением совершенно иного склада — язвительным, блестящим математиком из MIT, одним из отцов-основателей «символьного» подхода к ИИ. Он считал, что разум — это манипуляция символами по четким логическим правилам, а не какая-то аналоговая каша из подстраивающихся железок. Перцептрон с его мистическим «обучением» казался ему ересью и шарлатанством, угрожающим «настоящему» искусственному интеллекту.

Мински вместе с коллегой Сеймуром Пейпертом сели за

стол, взяли перцептрон и препарировали его с математической беспощадностью. В 1969 году они выпустили книгу с убийственным названием «Перцептроны» (Perceptrons). Это была блестящая и уничтожительная работа. Они математически строго доказали, что простой однослойный перцептрон, столь разрекламированный Розенблаттом, не может решить даже элементарную логическую функцию «Исключающее ИЛИ» (XOR). Проще говоря: если даны два утверждения, и правдой может быть только одно из них, но не оба сразу, машина Розенблатта в принципе не способна это понять. Она слепа. И добавление тысяч новых нейронов ничего не изменит. Фундамент оказался порочным.

Удар был смертельным. Книга Мински и Пейперта, подкрепленная их колоссальным академическим авторитетом, практически мгновенно убила интерес и финансирование всего направления. Военные закрыли проекты. Правительство перестало выделять гранты. Студентам рекомендовали не связываться с нейронными сетями, чтобы не губить карьеру.

Фрэнк Розенблатт не смог этого пережить. Он пытался бороться, разрабатывал более сложные модели, но его звезда закатилась. Он погиб в 1971 году при загадочных обстоятельствах — разбился на своей яхте в день своего 43-летия. Многие говорили о несчастном случае. Другие — о са-

моубийстве, сломленности человека, чью мечту растоптали.

Так закончилась первая великая эра нейросетей. Механическое дитя, подававшее такие надежды, было объявлено мертворожденным. Началась Первая зима искусственного интеллекта — долгий, мрачный период, длившийся почти полтора десятилетия. Немногие оставшиеся верные идее ученые, такие как Джеффри Хинтон, которого мы встретим позже, работали в полной безвестности, зная, что правда была не на стороне Мински... или, точнее, не совсем на его стороне. Мински и Пейперт доказали ограниченность лишь однослойных сетей, но их приговор на долгие годы ошибочно распространили на всю область. Они знали, что решение есть — многослойные сети. Но никто в мире еще не знал, как их обучить. Этот секрет будет храниться в тишине научных библиотек еще долгих семнадцать лет. А пока... на землю опустилась снежная тишина.

Глава 3. Долгая зима и одинокие хранители огня

Если предыдущая глава была похожа на яркую вспышку фейерверка, оборвавшуюся внезапной тьмой, то теперь представь себе долгую, глухую, бесснежную зиму. Без фанфар, без заголовков в газетах, без конференций с шампанским. Так выглядели 1970-е и начало 1980-х для тех, кто верил в нейронные сети. Это было время великого исхода. Термин «нейросеть» стал в приличном академическом обществе почти ругательством. Произнести его на кафедре значило вызвать снисходительную усмешку или откровенное презрение. «Этим же ещё Мински доказал...», — говорили авторитетные профессора, и спорить с ними было карьерным самоубийством.

Мейнстрим исследований искусственного интеллекта ушёл в совершенно другое русло — в то, что позже назовут «добрым старым ИИ» (GOFAI, Good Old-Fashioned AI). Это был подход Мински: интеллект как набор чётких, формальных правил. Экспертные системы. Если ты хочешь, чтобы машина ставила медицинский диагноз, ты должен вручную записать все правила: «ЕСЛИ у пациента температура

И сыпь И боль в горле, И НЕТ кашля, ТО, ВОЗМОЖНО, это скарлатина». Тысячи и тысячи таких правил, выуженных из голов живых врачей «инженерами по знаниям». Эти системы работали, иногда даже неплохо. Они могли управлять химическими заводами, находить месторождения полезных ископаемых. Они были логичны, прозрачны и понятны.

Но у них была фатальная слабость: они были абсолютно слепы и беспомощны перед всем, что не укладывалось в правила. Увидеть кошку на фотографии? Невозможно. Понять смысл произнесённой с акцентом фразы? Исключено. Мир бесконечно сложен и грязен, его не запихнуть в таблицу «ЕСЛИ-ТО». Реальный интеллект — это не свод законов. Это интуиция, это умение вычленять закономерности из хаоса. Именно это и обещали нейросети, но их могила, казалось, была залита бетоном.

И всё же в этом бетоне оставались крошечные трещины. По всему миру, в полном одиночестве, работали несколько упрямцев, которые знали, что Мински и Пейперт убили не всю идею. Они доказали ограниченность лишь однослойных сетей. А что, если слоёв будет много? Что, если информация пойдёт сквозь каскад нейронов, преобразуясь на каждом этапе? Теоретически это решало проблему XOR и открывало путь к настоящему обучению. Но как, чёрт возьми, обучать такую многослойную сеть? Как узнать, какой именно нейрон

в глубине этой «тёмной материи» ошибся, и насколько нужно подкрутить его веса? Сигнал об ошибке приходит только на самый последний слой, на выход сети. А что делать со слоями скрытыми? Эта проблема так и называлась — «проблема присвоения значимости» (credit assignment problem). Она казалась неразрешимой.

Одним из таких хранителей огня был Пол Вербос. В 1974 году он, будучи аспирантом Гарварда, в своей диссертации предложил решение. Оно было настолько изящным математически, что его современники просто не поняли, на что смотрят. Вербос предложил рассматривать нейронную сеть не как набор дискретных переключателей, а как сложную математическую функцию. И чтобы понять, как изменить любой вес в сети, нужно взять производную ошибки по этому весу. Это то же самое, что спросить: «Если я чуть-чуть шевельну вот этот рычажок, как сильно изменится итоговая ошибка?». И делать это не вручную, а по цепному правилу из школьного курса математики, последовательно прогоняя ошибку от выхода обратно ко входу. Отсюда и название — метод обратного распространения ошибки (backpropagation).

Работа Вербоса прошла почти незамеченной. Его научный руководитель не увидел в этом перспектив, диссертация пылилась на полке. Зима продолжалась.

Через несколько лет, в начале 1980-х, искра вспыхнула снова и уже громче. На этот раз в Японии. Кунихико Фукусима создал «Когнитрон», а затем «Неокогнитрон» — иерархическую многослойную сеть, способную распознавать рукописные символы. Его архитектура была вдохновлена устройством зрительной коры мозга: простые клетки реагируют на линии, сложные — на комбинации линий, и так далее. Это был прообраз будущих сверточных сетей. Но и его работа не смогла растопить лёд глобального скепсиса. Слишком силён был авторитет Мински, слишком свежа была рана от краха перцептрона.

Но история, как это часто бывает, любит точку бифуркации. И этой точкой стал тихий 1986 год. В журнале Nature выходит статья под скромным названием «Learning representations by back-propagating errors». Авторы: Дэвид Румельхарт, Джеффри Хинтон и Рональд Уильямс.

Это была бомба. Не потому что они придумали что-то принципиально новое — Вербос уже описал этот метод. Они сделали нечто гораздо более важное: они показали, что это работает. Причём работает на задачах, которые казались исключительной прерогативой человека. Например, их многослойная сеть училась формировать правильную форму прошедшего времени английских глаголов, включая исключе-

ния. Она получала на вход корень, на выходе должна была выдать прошедшую форму. Она не просто запоминала таблицу. Она улавливала правило, а потом, подобно ребёнку, начинала совершать «ошибки»: брала правильные глаголы и применяла к ним правило для неправильных («go» -> «goed»), пока не доучивала исключения. Это было ошеломляюще. Машина, которая учится как человек, ошибается как человек и исправляется как человек!

Джеффри Хинтон, праправнук логика Джорджа Буля, человек с несгибаемой волей и тихим голосом, стал знаменем возрождения. Он читал лекции, на которые снова начали приходить люди. Он убеждал, что символьный ИИ — это тупик. Интеллект не программируется, он выращивается. Мозг не начинает с чистого листа, полного правил, — он начинает с гибкой сети связей, которая лепится под давлением данных.

Зима начала отступать. Лёд треснул. Но до весеннего половодья и буйного цветения было ещё далеко. В распоряжении учёных всё ещё были смехотворно маленькие наборы данных и слабые процессоры. Однако в их руках наконец-то оказался правильный алгоритм — алгоритм, который умел учиться на своих ошибках, протаскивая сигнал обучения сквозь все слои искусственных нейронов. Механизм был запущен.

В следующей главе мы увидим, как ещё один «хранитель», француз Ян Лекун, возьмёт этот алгоритм, скрестит его с идеями Фукусимы и создаст сеть, которая действительно научится «видеть» цифры. Но и этого будет мало. Чтобы случился Большой взрыв, нужны будут две вещи: невероятно много данных и невероятно мощные вычислители. И они появятся из совершенно неожиданного источника: из мира видеоигр и из желания одного человека каталогизировать весь визуальный мир. Но это уже совсем другая история.

Глава 4. Свёртка. Как Ян Лекун научил машины видеть

К концу 1980-х годов метод обратного распространения ошибки перестал быть тайным знанием горстки отшельников. О нём говорили, его изучали, с ним экспериментировали. Но оставалась одна гигантская, унизительная проблема. Теоретически многослойные сети могли научиться чему угодно. На практике же они захлёбывались и тонули при малейшем усложнении задачи. Представь, что ты хочешь научить сеть распознавать простую рукописную цифру "3" на картинке размером 32 на 32 пикселя. Это крошечная чёрно-белая иконка. Но даже здесь требуется сеть, в которой каждый из 1024 пикселей-входов должен быть соединён с каждым нейроном первого скрытого слоя. Тысячи и тысячи связей. А если картинка чуть сдвинулась на пару пикселей влево? Для сети это совершенно другой набор входных сигналов. Она не понимает, что сдвинутая "тройка" — это всё та же "тройка". Ей нужно увидеть её во всех возможных положениях, на всех возможных фонах, под всеми возможными наклонами. Это требовало астрономического числа примеров и вычислительных мощностей, которых тогда просто не существовало.

Для решения этой проблемы нужен был кто-то, кто посмотрит на неё не как программист, а как нейробиолог. Этим человеком стал француз Ян Лекун. Высокий, худой, с неизменной трубкой и слегка рассеянным взглядом философа, он был очарован работами Фукусимы и его «Неокогнитрона». И его осенила идея, элегантная в своей биологической достоверности.

Как устроено зрение у млекопитающих? Эксперименты Хьюбела и Визеля ещё в 1960-х показали: в первичной зрительной коре есть так называемые «простые клетки». Каждая из них реагирует не на целую картинку, а на очень маленький, конкретный фрагмент поля зрения. Одна клетка возбуждается, только когда видит вертикальную линию в строго определённом месте. Другая — линию под углом 45 градусов. Эти клетки — детекторы элементарных признаков. Дальше информация передаётся «сложным клеткам», которые собирают сигналы от нескольких простых и уже не так чувствительны к точному положению признака. Увидев вертикальную линию чуть левее или чуть правее, они всё равно скажут: «Да, вертикальная линия где-то здесь». Этот принцип иерархии — от простого к сложному, от конкретного к абстрактному — и есть секрет нашего зрения.

Лекун решил реализовать это в кремнии. В 1989 году, ра-

ботая в Bell Labs, он предложил архитектуру, которую назвал свёрточной нейронной сетью (Convolutional Neural Network, CNN).

Идея была дерзкой и гениальной. Вместо того чтобы соединять каждый пиксель с каждым нейроном, Лекун использовал операцию свёртки. Представь себе маленькое «окошко» — фильтр размером 5 на 5 пикселей. Это и есть аналог «простой клетки». Это окошко скользит, как сканер, по всей картинке, и ищет свой единственный, простой признак — допустим, горизонтальный край. Везде, где оно находит этот край, на выходной карте загорается пиксель. Другой фильтр ищет вертикальный край. Третий — диагональный. Получается стопка «карт признаков». Затем в дело вступает пулинг (subsampling) — слой, который сжимает эти карты, усредняя сигналы от нескольких соседних нейронов. Это аналог «сложной клетки». Мы теряем точное местоположение признака, но приобретаем обобщение: «Где-то в этом квадрате был вертикальный край».

А затем этот процесс повторяется снова и снова. Следующий свёрточный слой ищет комбинации признаков: «вертикальный край рядом с горизонтальным — это угол!». Ещё выше — комбинации углов: «четыре угла и линии между ними — это прямоугольник!». И так, слой за слоем, сеть сама, автоматически, выстраивает всё более абстрактную иерар-

хию понятий. Ей больше не нужно видеть "тройку" во всех позициях. Благодаря пулингу, она становится инвариантной к небольшим сдвигам и искажениям. А благодаря весовому разделению (один и тот же фильтр используется для всей картинки) количество настраиваемых параметров драматически сокращается. Сеть становится глубже, но при этом проще для обучения.

Своё революционное детище Лекун назвал LeNet. И он нашёл для него идеальное, убойное применение. Не игрушечные квадратики, а суровая реальность капитализма.

В начале 1990-х американские банки и почтовая служба тонули в бумаге. Ежедневно через них проходили миллионы рукописных чеков и конвертов. Их нужно было сортировать, и для этого кому-то приходилось вручную вбивать написанные от руки цифры индексов и сумм. Армии клерков в офисах по всей стране занимались этим нудным, изматывающим трудом. Банки мечтали автоматизировать этот процесс и были готовы платить. Лекун и его команда обучили LeNet на тысячах образцов рукописных цифр. И сеть заработала. Причём заработала с такой точностью, что её можно было внедрять в промышленную эксплуатацию.

К концу 1990-х годов модификации LeNet, внедрённые в банкоматы и сортировочные машины, обрабатывали, по раз-

ным оценкам, от 10 до 20 процентов всех банковских чеков в Соединённых Штатах. Это были миллиарды долларов, сэкономленных на ручном труде. Свёрточная нейронная сеть, сама того не ведая, стала одним из крупнейших «белых воротничков» в финансовой системе. Но парадокс в том, что мир не кричал о революции. Никто не выносил Лекуна на обложки. Всё это работало тихо, под капотом, как какой-нибудь неприметный, но очень полезный контроллер. Это была победа, но победа локальная. Триумф нейросетей, загнанный в рамки одной конкретной, утилитарной задачи. Для того чтобы CNN, да и все остальные сети, совершили тот самый прыжок, который мы называем «Большим взрывом», нужны были ещё два компонента. Первый — океан данных. Не тысячи, а миллионы и миллионы размеченных примеров. И второй — чудовищная, немыслимая по тем временам вычислительная мощь.

Ирония судьбы в том, что и то, и другое человечество создало во многом ради развлечения. Данные породил интернет, заполнив фотографиями, текстами и видео каждый его уголок. А мощь пришла из индустрии компьютерных игр. Видеокарты (GPU), созданные для того, чтобы просчитывать миллионы полигонов в шутерах от первого лица, оказались идеальными машинами для расчёта миллионов нейронных связей. Оставался последний шаг: нужен был человек, который объединит эти три компонента — глубокую ар-

хитектуру, гигантский датасет и мощь GPU — в одном эксперименте. Этот человек уже учился в университете, и его звали Алекс Крижевский. Но прежде чем он вышел на сцену, одна блестящая женщина-учёный должна была создать тот самый «Эверест», который он покорит. Её звали Фей-Фей Ли.

Глава 5. Глаза для машин: Фей-Фей Ли, ImageNet и дар геймеров

Представь себе ребёнка, который учится говорить и распознавать предметы. Он не получает словарного определения каждой вещи с первого раза. Вместо этого родители и весь окружающий мир терпеливо, год за годом, показывают ему бесчисленное множество примеров. «Это кошка. И это кошка. А это, видишь, уже собака. Нет, это всё ещё кошка, просто пушистая и лежит в коробке». К тому моменту, когда ребёнок идёт в детский сад, его мозг обработал астрономическое количество визуальных образов, каждый из которых был снабжён ярлыком: словом, звуком, тактильным ощущением. Именно эта гигантская, размеченная реальностью выборка и создаёт фундамент человеческого интеллекта.

До середины 2000-х годов у компьютерного зрения такой выборки не было. Алгоритмы мучили на крошечных, игрушечных наборах данных. Самый популярный из них, Caltech-101, содержал чуть более 9000 изображений, разбитых на 101 категорию. Тысячи изображений! Этого катастрофически мало, чтобы охватить бесконечное разнообразие реального мира — ракурсов, освещения, окрасок, фона.

Сети, обученные на такой скудной диете, были близорукими и хрупкими. Они ломались, столкнувшись с реальной фотографией, сделанной под непривычным углом. Нужен был прорыв в данных. И этот прорыв совершила женщина, которая изначально вовсе не собиралась заниматься компьютерами.

Фей-Фей Ли родилась в Пекине, а в 16 лет переехала с семьёй в США. Её родители, принадлежавшие в Китае к интеллигенции, в Америке работали официантами и продавцами, чтобы дать дочери образование. Она хотела стать физиком, увлекалась тайнами Вселенной. Но однажды, наблюдая за своим ребёнком, она задумалась о чуде человеческого восприятия. Малыш смотрел на мир широко раскрытыми глазами и, казалось, впитывал его без всяких усилий. Как? Как из потока фотонов, падающих на сетчатку, рождается понимание «это стул, на него можно сесть»? Этот вопрос заворожил её и привёл в область компьютерного зрения.

К середине 2000-х Фей-Фей Ли была уже профессором. И она выдвинула идею, которая поначалу показалась коллегам безумием, прожектёрством, достойным Дон Кихота. Она предложила создать базу данных из миллионов изображений, охватывающую не сотни, а тысячи и десятки тысяч категорий объектов. Всех. Не только «кошки» и «собаки», но и «каноэ», «скрипки», «закаты», «улыбки», «текстуры ржаво-

го металла» — всё богатство видимого мира. Причём каждое изображение должно быть размечено вручную живым человеком.

Её отговаривали. Гранты не давали. «Вы с ума сошли, — говорили маститые профессора. — Вам понадобится целая армия дешёвой рабочей силы, годы труда и миллионы долларов. Это технически невозможно». Но Фей-Фей Ли верила в свою идею с упорством, которое бывает только у людей, пересёкших океан и начавших жизнь с нуля. Она понимала то, что позже станет аксиомой: данные — это топливо для алгоритмов. Без топлива самый совершенный двигатель — просто груда металла.

И здесь случилось чудо, имя которому — краудсорсинг. Как раз в это время набирал силу Amazon Mechanical Turk — онлайн-платформа, позволявшая нанимать сотни тысяч людей по всему миру для выполнения крошечных, простых заданий за копейки. Это и была та самая армия. Фей-Фей Ли и её команда наняли этих «цифровых рабочих» и поставили перед ними простую задачу: смотреть на картинки и говорить, что на них изображено.

Процесс был титаническим. Нужно было не просто собрать фотографии из поисковиков. Их требовалось очистить, проверить, привести к единому стандарту. А затем —

разметить. Не один раз, а многократно, для перепроверки. Десятки тысяч анонимных людей в разных часовых поясах годами сидели перед мониторами и кликали: «автомобиль», «дерево», «лицо», «подушка», «водопад». Так, крупица за крупицей, рождалась ImageNet.

К 2009 году работа была завершена. Результат превзошёл самые смелые ожидания. ImageNet содержал более 14 миллионов изображений, аккуратно рассортированных по 21 841 категории. Это был Эверест данных. Цифровой Ноев ковчег, в котором имелось по паре каждого визуального существа и явления. Но, как это часто бывает с великими творениями, поначалу его никто не оценил. На ведущей конференции по компьютерному зрению CVPR доклад Фей-Фей Ли приняли, но выделили ей лишь скромную постерную сессию, а не устный доклад. Эверест стоял, но желающих его штурмовать пока не находилось.

Нужно было превратить каталог в соревнование. Фей-Фей Ли и её коллеги объявили о запуске ImageNet Large Scale Visual Recognition Challenge (ILSVRC) — ежегодного конкурса, где лучшие алгоритмы мира будут сражаться за точность распознавания на её гигантском датасете. И чтобы подогреть интерес, они выделили подмножество из 1,2 миллиона изображений и 1000 тщательно отобранных категорий. Это был вызов. Приз, манивший лучших гладиаторов мира

ИИ.

Оставалась одна проблема: вычислительная мощь. CNN Лекуна были хороши, но на ImageNet они захлебнулись бы. Обучать глубокие сети на миллионах картинок на обычных процессорах (CPU) было всё равно что пытаться выкопать Панамский канал лопатой. Теоретически можно, но займёт годы или десятилетия. И тут своё слово сказала индустрия, которая меньше всего думала о науке. Индустрия компьютерных игр.

Пока учёные бились над распознаванием кошек, геймеры по всему миру требовали всё более реалистичной графики. Чтобы рисовать взрывы, тени и отражения в реальном времени, процессорам не хватало духу. Так появились специализированные графические процессоры (GPU) — видеокарты. Их архитектура была принципиально иной. Если CPU — это горстка очень сильных, универсальных математиков, решающих сложные уравнения последовательно, то GPU — это армия из тысяч слабых, но усердных клерков, которые могут одновременно складывать простые числа. А нейросеть, особенно CNN, в своей основе — это именно что миллиарды однотипных простых операций умножения и сложения матриц. Задача, идеально параллелящаяся на тысячи ядер видеокарты.

Конец ознакомительного фрагмента.