

Оглавление

Пролог: Испуганные 11

Часть I

Предыстория

Глава 1. Истоки искусственного интеллекта29

Глава 2. Нейронные сети и подъем машинного обучения 50

Глава 3. Весна ИИ59

Часть II

Смотреть и видеть

Глава 4. Что, где, когда, кто и зачем?87

Глава 5. Сверточные нейронные сети и *ImageNet*104

Глава 6. Подробнее об обучаемых машинах123

Глава 7. Надежный и этичный ИИ149

Часть III

Обучение игре

Глава 8. Награды для роботов165

Глава 9. Игра началась179

Глава 10. Не ограничиваясь играми204

Часть IV
Искусственный интеллект
и естественный язык

Глава 11. Слова и их соседи	217
Глава 12. Перевод как кодирование и декодирование	240
Глава 13. Спроси меня о чем угодно	259

Часть V
Барьер понимания

Глава 14. О понимании	282
Глава 15. Знания, абстракции и аналогии в искусственном интеллекте ..	296
Глава 16. Вопросы, ответы и гипотезы	319

<i>Благодарности</i>	339
<i>Примечания</i>	341
<i>Сведения об иллюстрациях</i>	371
<i>Предметно-именной указатель</i>	375

Часть I
Предыстория

Глава 1

Истоки искусственного интеллекта

Два месяца и десять мужчин в Дартмуте

Мечта о создании разумной машины — машины, которая разумна в той же степени, что и человек, или даже превосходит его, — появилась много столетий назад, но стала частью современной науки с началом эры цифровых компьютеров. Идеи, приведшие к созданию первых программируемых компьютеров, родились из стремления математиков понять человеческое мышление — в частности, логику — как механический процесс “манипуляции символами”. Цифровые компьютеры, по сути, представляют собой символьные манипуляторы, которые оперируют комбинациями символов 0 и 1. Пионеры вычислительной техники, включая Алана Тьюринга и Джона фон Неймана, проводили аналогии между работой компьютеров и мозга человека, и им казалось очевидным, что человеческий разум можно воссоздать в компьютерной программе.

Большинство людей из сферы искусственного интеллекта считают, что официально она выделилась в отдельную дисциплину на маленьком семинаре, который был организован молодым математиком Джоном Маккарти и состоялся в 1956 году в Дартмуте.

В 1955 году Маккарти, которому тогда было двадцать восемь лет, поступил преподавателем на математический факультет Дартмутского колледжа. В студенческие годы он изу-

чал психологию и новую “теорию автоматов” (которая впоследствии стала информатикой) и заинтересовался идеей о создании думающей машины. В аспирантуре на кафедре математики Принстонского университета Маккарти познакомился с Марвином Минским, который учился вместе с ним и тоже интересовался потенциалом разумных компьютеров. Окончив аспирантуру, Маккарти некоторое время работал в Лабораториях Белла и ИВМ, где сотрудничал с основателем теории информации Клодом Шенноном и пионером электротехники Натаниэлем Рочестером. Оказавшись в Дартмуте, Маккарти убедил Минского, Шеннона и Рочестера помочь ему с организацией “двухмесячного семинара по изучению искусственного интеллекта, который планировалось провести летом 1956 года с участием десяти человек”¹. Термин “искусственный интеллект” был предложен Маккарти, который хотел отделить эту сферу от связанной с ней кибернетики². Позже Маккарти признал, что название никому не нравилось, ведь целью был *настоящий*, а не “искусственный” интеллект, но “без названия было не обойтись”, поэтому он стал использовать понятие “искусственный интеллект”³.

Четыре организатора летнего семинара подали заявку на получение финансирования в Фонд Рокфеллера. Они написали, что в основе планируемого исследования лежит “предположение, что каждый аспект обучения и любую другую характеристику интеллекта теоретически можно описать так точно, что можно создать машину для его моделирования”⁴. В заявке перечислялись основные темы для обсуждения — обработка естественного языка, нейронные сети, машинное обучение, абстрактные концепции и рассуждения, творческие способности, — и они до сих пор определяют сферу искусственного интеллекта.

Хотя самые мощные компьютеры в 1956 году были примерно в миллион раз медленнее современных смартфонов, Маккарти с коллегами полагали, что создание ИИ не за горами: “Мы считаем, что значительного прогресса по одной или нескольким из этих задач можно добиться, если группа

правильно подобранных ученых получит возможность поработать над ними вместе в течение лета”⁵.

Вскоре возникли трудности, знакомые любому, кто хоть раз пытался организовать научный семинар. Фонд Рокфеллера согласился предоставить лишь половину запрашиваемой суммы, а убедить участников приехать и остаться оказалось гораздо сложнее, чем Маккарти мог предположить. О соглашения между ними не шло и речи. Было много интересных дискуссий и много противоречий. Как обычно бывает на подобных встречах, “у каждого была своя идея, большое сомнение и горячий интерес к собственному плану”⁶. Тем не менее дартмутское лето ИИ помогло добиться очень важных результатов. У области исследований появилось название. Были очерчены ее общие цели. Будущая “большая четверка” пионеров сферы — Маккарти, Минский, Аллен Ньюэлл и Герберт Саймон — встретились и построили ряд планов на будущее. По какой-то причине эти четверо после семинара смотрели в будущее искусственного интеллекта с оптимизмом. В начале 1960-х годов Маккарти основал Стэнфордский проект в области искусственного интеллекта с “целью за десять лет сконструировать полностью разумную машину”⁷. Примерно в то же время будущий нобелевский лауреат Герберт Саймон предсказал: “В ближайшие двадцать лет машины смогут выполнять любую работу, которая под силу человеку”⁸. Вскоре после этого основатель Лаборатории ИИ в MIT Марвин Минский дал прогноз, что “до смены поколений... задачи создания «искусственного интеллекта» будут в основном решены”⁹.

Основные понятия и работа с ними

Пока не произошло ни одно из предсказанных событий. Насколько далеки мы от цели сконструировать “полностью разумную машину”? Потребуется ли нам для этого осуществить обратное проектирование сложнейшего человеческого мозга

или же мы найдем короткий путь, где хитрого набора пока неизвестных алгоритмов будет достаточно, чтобы создать полностью разумную в нашем представлении машину? Что вообще имеется в виду под “полностью разумной машиной”?

“Определяйте понятия... или мы никогда не поймем друг друга”¹⁰. Этот призыв философа XVIII века Вольтера остается слабым местом дискуссий об искусственном интеллекте, поскольку главное понятие — интеллект — все еще не имеет четкого определения. Марвин Минский называл такие понятия, как “интеллект”, “мышление”, “познание”, “сознание” и “эмоция”, “словами-чемоданами”¹¹. Каждое из них, подобно чемодану, набито кучей разных смыслов. Понятие “искусственный интеллект” наследует эту проблему, в разных контекстах означая разное.

Большинство людей согласится, что человек наделен интеллектом, а пылинка — нет. Кроме того, мы в массе своей считаем, что человек разумнее червя. Коэффициент человеческого интеллекта измеряется по единой шкале, но мы также выделяем разные аспекты интеллекта: эмоциональный, вербальный, пространственный, логический, художественный, социальный и так далее. Таким образом, интеллект может быть бинарным (человек либо умен, либо нет), может иметь диапазон (один объект может быть разумнее другого), а может быть многомерным (один человек может обладать высоким вербальным, но низким эмоциональным интеллектом). Понятие “интеллект” действительно напоминает набитый до отказа чемодан, который грозит лопнуть.

Как бы то ни было, сфера ИИ практически не принимает в расчет эти многочисленные различия. Усилия ученых направлены на решение двух задач: научной и практической. В научном направлении исследователи ИИ изучают механизмы “естественного” (то есть биологического) интеллекта, пытаясь внедрить его в компьютеры. В практическом направлении поборники ИИ просто хотят создать компьютерные программы, которые справляются с задачами не хуже или лучше людей, и не заботятся о том, *думают* ли эти про-

граммы таким же образом, как люди. Если спросить исследователей ИИ, какие цели — научные или практические — они преследуют, многие в шутку ответят, что это зависит от того, кто в данный момент финансирует их работу.

В недавнем отчете о текущем состоянии ИИ комитет видных исследователей определил эту область как “раздел информатики, изучающий свойства интеллекта посредством синтеза интеллекта”¹². Да, определение закольцовано. Впрочем, тот же комитет признал, что определить область сложно, но это и к лучшему: “Отсутствие точного, универсального определения ИИ, вероятно, помогает области все быстрее расти, развиваться и совершенствоваться”¹³. Более того, комитет отмечает: “Практики, исследователи и разработчики ИИ ориентируются по наитию и руководствуются принципом «бери и делай»”.

Анархия методов

Участники Дартмутского семинара 1956 года озвучивали разные мнения о правильном подходе к разработке ИИ. Одни — в основном математики — считали, что языком рационального мышления следует считать математическую логику и дедуктивный метод. Другие выступали за использование индуктивного метода, в рамках которого программы извлекают статистические сведения из данных и используют вероятности при работе с неопределенностью. Третьи твердо верили, что нужно черпать вдохновение из биологии и психологии и создавать программы по модели мозга. Как ни странно, споры между сторонниками разных подходов не утихают по сей день. Для каждого подхода было разработано собственное множество принципов и техник, поддерживаемых отраслевыми конференциями и журналами, но узкие специальности почти не взаимодействуют между собой. В недавнем исследовании ИИ отмечается: “Поскольку мы не имеем глубокого понимания интеллекта и не знаем, как создать общий ИИ, чтобы идти по пути настоящего прогресса, нам нужно

не закрывать некоторые направления исследований, а принимать «анархию методов», царящую в сфере ИИ»¹⁴.

Однако с 2010-х годов одно семейство ИИ-методов, в совокупности именуемых глубоким обучением (или глубокими нейронными сетями), выделилось из анархии и стало господствующей парадигмой ИИ. Многие популярные медиа сегодня ставят знак равенства между понятиями “искусственный интеллект” и “глубокое обучение”. При этом они совершают досадную ошибку, и мне стоит прояснить различия между терминами. ИИ — это область, включающая широкий спектр подходов, цель которых заключается в создании наделенных интеллектом машин. Глубокое обучение — лишь один из этих подходов. Глубокое обучение — лишь один из множества методов в области “машинного обучения”, подобласти ИИ, где машины “учатся” на основе данных или собственного “опыта”. Чтобы лучше понять эти различия, нужно разобраться в философском расколе, который произошел на заре исследования ИИ, когда произошло разделение так называемых символического и субсимволического ИИ.

Символический ИИ

Давайте сначала рассмотрим *символический* ИИ. Программа символического ИИ знает слова или фразы (“символы”), как правило понятные человеку, а также правила комбинирования и обработки этих символов для выполнения поставленной перед ней задачи.

Приведу пример. Одной ранней программе ИИ присвоили громкое имя “Универсальный решатель задач” (*General Problem Solver*, или GPS)¹⁵. (Прошу прощения за сбивающую с толку аббревиатуру: Универсальный решатель задач появился раньше системы глобального позиционирования, ныне известной как GPS.) УРЗ мог решать такие задачи, как задача о миссионерах и людоедах, над которой вы, возможно, ломали голову в детстве. В этой известной задаче три миссио-

нера и три людоеда должны переправиться через реку на лодке, способной выдержать не более двух человек. Если на одном берегу окажется больше (голодных) людоедов, чем (аппетитных) миссионеров, то... думаю, вы поняли, что произойдет. Как всем шестерым переправиться на другой берег без потерь?

Создатели УРЗ, когнитивисты Герберт Саймон и Аллен Ньюэлл, записали, как несколько студентов “размышляют вслух”, решая эту и другие логические задачи. Затем Саймон и Ньюэлл сконструировали программу таким образом, чтобы она копировала ход рассуждений студентов, который ученые признали их мыслительным процессом.

Я не буду подробно описывать механизм работы УРЗ, но его символическую природу можно разглядеть в формулировке программных инструкций. Чтобы поставить задачу, человек писал для УРЗ подобный код:

ТЕКУЩЕЕ СОСТОЯНИЕ:

ЛЕВЫЙ-БЕРЕГ = [3 МИССИОНЕРА,

3 ЛЮДОЕДА, 1 ЛОДКА]

ПРАВЫЙ-БЕРЕГ = [ПУСТО]

ЖЕЛАЕМОЕ СОСТОЯНИЕ:

ЛЕВЫЙ-БЕРЕГ = [ПУСТО]

ПРАВЫЙ-БЕРЕГ = [3 МИССИОНЕРА,

3 ЛЮДОЕДА, 1 ЛОДКА]

Если говорить обычным языком, эта инструкция показывает, что изначально левый берег реки “содержит” трех миссионеров, трех людоедов и одну лодку, в то время как правый не содержит ничего. Желаемое состояние определяет цель программы — переправить всех на правый берег реки.

На каждом шаге программы УРЗ пытается изменить текущее состояние, чтобы сделать его более похожим на желаемое состояние. В этом коде у программы есть “операторы” (в форме подпрограмм), которые могут преобразовывать текущее состояние в новое состояние, и “правила”, кодирующие ограничения

задачи. Например, один оператор перемещает некоторое количество миссионеров и людоедов с одного берега реки на другой:

ПЕРЕМЕСТИТЬ (#МИССИОНЕРОВ, #ЛЮДОЕДОВ,
С-БЕРЕГА, НА-БЕРЕГ)

Слова в скобках называются параметрами, и после запуска программа заменяет эти слова на числа или другие слова. Параметр #МИССИОНЕРОВ заменяется на количество перемещаемых миссионеров, параметр #ЛЮДОЕДОВ заменяется на количество перемещаемых людоедов, а параметры С-БЕРЕГА и НА-БЕРЕГ заменяются на “ЛЕВЫЙ-БЕРЕГ” и “ПРАВЫЙ-БЕРЕГ” в зависимости от того, с какого берега нужно переместить миссионеров и людоедов. В программе закодировано знание, что лодка перемещается вместе с миссионерами и людоедами.

Прежде чем программа сумеет применить этот оператор с конкретными значениями параметров, она должна свериться с закодированными правилами: так, за один раз можно переместить не более двух человек, а если в результате применения оператора на одном берегу окажется больше людоедов, чем миссионеров, применять его нельзя.

Хотя эти символы обозначают знакомые людям понятия — “миссионеры”, “людоеды”, “лодка”, “левый берег”, — запускающий программу компьютер, конечно, не понимает значения символов. Можно заменить параметр “МИССИОНЕРЫ” на “z372v” или любой другой бессмысленный набор знаков, и программа будет работать точно так же. Отчасти поэтому она называется *Универсальным* решателем задач. Компьютер определяет “значение” символов в зависимости от того, как их можно комбинировать и соотносить друг с другом и как ими можно оперировать.

Адвокаты символического подхода к ИИ утверждали, что невозможно наделить компьютер разумом, не написав программы, копирующие человеческий мозг. Они полагали, что для создания общего интеллекта необходима лишь верная про-

грамма обработки символов. Да, эта программа работала бы гораздо сложнее, чем в примере с миссионерами и людоедами, но все равно состояла бы из символов, комбинаций символов, правил и операций с символами. В итоге символический ИИ, примером которого стал Универсальный решатель задач, доминировал в сфере ИИ в первые три десятилетия ее существования. Самым заметным его воплощением стали *экспертные системы*, которые использовали созданные людьми правила для компьютерных программ, чтобы выполнять такие задачи, как постановка медицинских диагнозов и принятие юридических решений. Символический ИИ по-прежнему применяется в нескольких сферах — я приведу примеры таких систем позже, в частности при обсуждении подходов ИИ к построению логических выводов и здравому смыслу.

Субсимволический ИИ: перцептроны

Вдохновением для символического ИИ послужила математическая логика и описание людьми своих сознательных мыслительных процессов. *Субсимволический* ИИ, напротив, вдохновлялся нейробиологией и стремился ухватить порой бессознательные мыслительные процессы, лежащие в основе так называемых процессов быстрого восприятия, например распознавания лиц или произносимых слов. Программы субсимволического ИИ не содержат понятного людям языка, который мы наблюдали в примере с миссионерами и людоедами. По сути, они представляют собой набор уравнений — настоящие дебри непонятных операций с числами. Как я поясню чуть дальше, такие системы на основе данных учатся выполнять поставленную перед ними задачу.

Одной из первых субсимволических ИИ-программ, созданных по модели мозга, стал перцептрон, изобретенный в конце 1950-х годов психологом Фрэнком Розенблаттом¹⁶. Сегодня термин “перцептрон” кажется заимствованным из научной фантастики пятидесятых годов (как мы увидим, вскоре

за ним последовали “когнитрон” и “неокогнитрон”), но перцептрон стал важной вехой развития ИИ и может считаться авторитетным прадедом самого успешного инструмента современного ИИ, глубоких нейронных сетей.

Розенблатт изобрел перцептрон, обратив внимание на то, как нейроны обрабатывают информацию. Нейрон — это клетка мозга, которая получает электрический или химический импульс от связанных с нею нейронов. Грубо говоря, нейрон суммирует все импульсы, которые получает от других нейронов, и сам посылает импульс, если итоговая сумма превышает определенный порог. Важно, что разные связи (*синапсы*) конкретного нейрона с другими нейронами имеют разную силу, а потому, суммируя импульсы, нейрон придает больше веса импульсам от сильных связей, чем импульсам от слабых связей. Нейробиологи полагают, что поправки на силу связей между нейронами — важнейший элемент процесса обучения, происходящего в мозге.

С точки зрения специалиста по информатике (или, как в случае с Розенблаттом, психолога), обработку информации нейронами можно смоделировать в компьютерной программе — перцептроне, — которая преобразует много численных входных сигналов в один выходной сигнал. Аналогия между нейроном и перцептроном показана на рис. 1. На рис. 1А мы видим нейрон с ветвистыми дендритами (волоконками, которые проводят входящие импульсы в клетку), телом клетки и аксоном (или выводным каналом). На рис. 1В изображен простой перцептрон. Как и нейрон, перцептрон суммирует все входящие сигналы. Если итоговая сумма равняется *порогу* перцептрона или превышает его, перцептрон выдает значение 1 (“передает сигнал”); в противном случае он выдает значение 0 (“не передает сигнал”). Чтобы смоделировать различную силу связей нейрона, Розенблатт предложил присваивать каждому входному сигналу перцептрона численный *вес* и умножать входной сигнал на его вес, прежде чем прибавлять к сумме. *Порог* перцептрона — это число, определяемое программистом (или, как мы увидим, узнаваемое самим перцептроном).

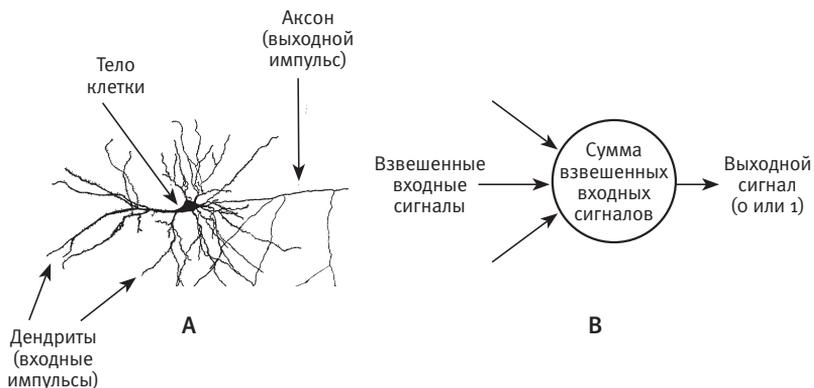


Рис. 1. А: нейрон в мозге; В: простой перцептон

Иными словами, перцептрон — это простая программа, которая принимает решение “да или нет” (1 или 0) в зависимости от того, достигает ли сумма взвешенных входных сигналов порогового значения. Вероятно, вы тоже время от времени принимаете такие решения в жизни. Например, вы узнаете мнение нескольких друзей о конкретном фильме, но вкусам одних друзей доверяете больше, чем вкусам других. Если сумма “дружеских восторгов” — при большем весе мнений тех друзей, которым вы доверяете больше, — достаточно высока (то есть превышает некоторый неосознанный порог), вы решаете посмотреть фильм. Именно так перцептрон выбирал бы фильмы к просмотру, если бы у него были друзья.