Часть I

Искусственный интеллект: его истоки и проблемы

Всему есть начало, как говорил Санчо Панса, и это начало должно опираться на нечто, ему предшествующее. Индусы придумали слона, который удерживал мир, но им пришлось поставить его на черепаху. Нужно отметить, что изобретение состоит в сотворении не из пустоты, но из хаоса: в первую очередь следует позаботиться о материале...

— Мэри Шелли (Mary Shelley), *Франкенштейн*

Попытка дать определение искусственному интеллекту

Искусственный интеллект (ИИ) можно определить как область компьютерной науки, занимающуюся автоматизацией разумного поведения. Это определение наиболее точно соответствует содержанию данной книги, поскольку в ней ИИ рассматривается как часть компьютерной науки, которая опирается на ее теоретические и прикладные принципы. Эти принципы сводятся к структурам данных, используемым для представления знаний, алгоритмам применения этих знаний, а также языкам и методикам программирования, используемым при их реализации.

Тем не менее это определение имеет существенный недостаток, поскольку само понятие интеллекта не очень понятно и четко сформулировано. Большинство из нас уверены, что смогут отличить “разумное поведение”, когда с ним столкнутся. Однако вряд ли кто-нибудь сможет дать интеллекту определение, достаточно конкретное для оценки предположительно разумной компьютерной программы и одновременно отражающее жизнеспособность и сложность человеческого разума.

Итак, проблема определения *искусственного* интеллекта сводится к проблеме определения интеллекта вообще: является ли он чем-то единым, или же этот термин объединяет набор разрозненных способностей? В какой мере интеллект можно создать, а в ка-

кой он существует априори? Что именно происходит при таком создании? Что такое творчество? Что такое интуиция? Можно ли судить о наличии интеллекта только по наблюдаемому поведению, или же требуется свидетельство наличия некоего скрытого механизма? Как представляются знания в нервных тканях живых существ, и как можно применить это в проектировании интеллектуальных устройств? Что такое самоанализ и как он связан с разумностью? И, более того, необходимо ли создавать интеллектуальную компьютерную программу по образу и подобию человеческого разума, или же достаточно строго "инженерного" подхода? Возможно ли вообще достичь разумности посредством компьютерной техники, или же сущность интеллекта требует богатства чувств и опыта, присущего лишь биологическим существам?

На эти вопросы ответа пока не найдено, но все они помогли сформировать задачи и методологию, составляющие основу современного ИИ. Отчасти привлекательность искусственного интеллекта в том и состоит, что он является оригинальным и мощным орудием для исследования именно этих проблем. ИИ предоставляет средство и испытательную модель для теорий интеллекта: такие теории могут быть переформулированы на языке компьютерных программ, а затем испытаны при их выполнении.

По этим причинам наше первоначальное определение, очевидно, не дает однозначной характеристики для этой области науки. Оно лишь ставит новые вопросы и открывает парадоксы в области, одной из главных задач которой является поиск самоопределения. Однако проблема поиска точного определения ИИ вполне объяснима. Изучение искусственного интеллекта — еще молодая дисциплина, и ее структура, круг вопросов и методики не так четко определены, как в более зрелых науках, например, физике.

Искусственный интеллект призван расширить возможности компьютерных наук, а не определить их границы. Одной из важных задач, стоящих перед исследователями, является поддержание этих усилий ясными теоретическими принципами.

Из-за специфики проблем и целей искусственный интеллект не поддается простому определению. Поэтому на первых порах просто опишем его как *спектр проблем и методологий, изучаемых разработчиками систем искусственного интеллекта*. Это определение может показаться глупым и бессмысленным, но оно отражает важный факт: искусственный интеллект, как и любая наука, является сферой интересов человека, и лучше всего рассматривать его в этом контексте.

Любая наука, включая ИИ, рассматривает некоторый круг проблем и разрабатывает подходы к их решению. Краткое изложение истории искусственного интеллекта, рассказ о личностях и их гипотезах, положенных в основу этой науки, поясняет, почему некоторые проблемы стали доминировать в этой области и почему для их решения были взяты на вооружение методы, описываемые в этой книге.

1

Искусственный интеллект: история развития и области приложения

Слушайте далее и вы еще более изумитесь ремеслам и богатствам природы, открытым мною. Величайшим было такое: в старину, если человек заболевал, у него не было защиты против болезни, ни исцеляющей еды, ни питья, ни мази; люди вымирали от отсутствия лекарств, но я показал им, как смешивать мягкие ингредиенты, чтобы изгонять всяческие хвори...

Это я сделал видимыми для человеческих очей пылающие знаки в небесах, что до тех пор были в тумане. Недра земли, скрытое благословение человечества, медь, железо, серебро и золото — осмелится ли кто-нибудь заявить, что он открыл их ранее меня? Я уверен, никто, если он не лжец. Говоря кратко: все ремесла, что есть у смертных, идут от Прометея.

— Эсхил (Aeschylus), *Прикованный Прометей*

1.1. Отношение к интеллекту, знанию и человеческому мастерству

Прометей говорит о результатах своего неповиновения богам Олимпа: его целью было не только украсть огонь для людей, но и просветить их посредством дара ума, *nous*, или же "сообразительности". Интеллект является основой всех разработанных человеком технологий и цивилизации вообще. Работа классического греческого драматурга иллюстрирует глубокую и давнюю уверенность в необычайной силе знания. Искусственный интеллект применяется во всех сферах наследия Прометея: медицине, психологии, биологии, астрономии, геологии и многих областях науки, которые Эсхил не в силах был себе представить.

Хотя поступок Прометея освободил людей от невежества, он навлек на него гнев Зевса. За кражу знаний, прежде принадлежавших лишь богам Олимпа Зевс приказал приковать Прометея к голой скале, чтобы стихии причиняли ему вечные страдания. Мысль о том, что человеческое стремление к знаниям является проступком перед богами или природой, прочно укоренилась в западной философии. На ней основана история Эдема,

она пронизывает сочинения Данте и Мильтона. И Шекспир, и древнегреческие трагики считали амбиции разума причинами всех бедствий. Упорная вера в то, что жажда знаний в конечном счете приведет к катастрофе, пережила и эпоху Возрождения, и век Просвещения и даже научные и философские открытия XIX и XX веков. Поэтому не стоит удивляться тому, что в научных и общественных кругах не утихает бурная полемика по поводу искусственного интеллекта.

И вправду, современная технология не развеяла древний страх губительных последствий интеллектуального честолюбия, она, скорее, сделала их более вероятными, а может, и неотвратимыми. Легенды о Прометее, Еве, Фаусте пересказываются на языке технологического общества. В своем предисловии к работе "Франкенштейн" (которая, кстати, носит подзаголовок "Современный Прометей") Мэри Шелли пишет:

> "Я была верным молчаливым слушателем долгих бесед между лордом Байроном и Шелли. В одной из них обсуждались различные философские доктрины, в частности, сущность первопричин жизни, возможность их постижения и изучения. Они говорили об экспериментах доктора Дарвина (я имею в виду не то, что действительно делал доктор, а то, что ему приписывали), который хранил вермишель в стеклянной емкости, пока она не начала сама двигаться каким-то непостижимым образом. Это не значит, что таким образом можно дать жизнь. Но, должно быть, возможно оживить труп. Об этом свидетельствует гальванизм: может быть, и можно изготовить составные части создания, соединить их вместе и наполнить живительным теплом". [Buttler, 1998]

Шелли демонстрирует нам, в какой мере научные достижения, такие как работы Дарвина и открытие электричества, убедили даже далеких от науки людей в том, что творения природы не являются божественной тайной — их можно "разбирать" и систематически изучать. Чудовище Франкенштейна — не продукт шаманских заклинаний или сделок с преисподней; его собрали из отдельно "изготовленных" компонентов и наполнили живительной силой электричества. Хотя наука девятнадцатого века не способна была понять цель изучения принципов и создания в полной мере разумного агента, она признавала мысль, что тайны жизни и разума можно приоткрыть с помощью научного анализа.

1.1.1. Историческая подоплека

К тому времени как Мэри Шелли окончательно и, вероятно, бесповоротно соединила современную науку с мифом о Прометее, философские корни современных работ в сфере искусственного интеллекта развивались уже несколько тысячелетий. Хотя моральные и культурные проблемы, поднятые искусственным интеллектом, интересны и важны, данное введение в б*о*льшей степени касается интеллектуального наследия ИИ. Логической отправной точкой этой истории можно считать гений Аристотеля, или, как его называл Данте, "мастера тех, кто знает". Аристотель объединил интуитивное понимание, тайны и предчувствия ранней греческой традиции с тщательным анализом и строгим мышлением, которому суждено было стать стандартом для современной науки.

Для Аристотеля наиболее пленительным аспектом природы была ее изменчивость. В работе "Физика" он определил свою "философию природы" как "изучение изменяющихся вещей". Он делал различие между *материей* и *формой*: например, скульптура сделана из *материи* бронзы и имеет *форму* человека. Изменение происходит в тот момент, когда бронзе придают другую форму. Разделение материи и формы представляет философский базис для современных научных концепций, таких как символьное исчисление или абстракция данных. В любом исчислении (даже в работе с числами!) мы манипулируем об-

разами, которые являются формой электромагнитной материи, а изменения формы этой материи передают аспекты процесса решения. Абстрагирование формы от средства ее представления не только позволяет производить вычисления над этой формой, но и служит основой теории структур данных — ядра современных компьютерных наук.

В своей работе "Метафизика" Аристотель разработал теории неизменных вещей — космологию и теологию. Но ближе всего к искусственному интеллекту подходит аристотелевская эпистемология, или наука познания, обсуждаемая в его "Логике". Аристотель считал эту книгу важным инструментом познания, поскольку чувствовал, что основой знания является изучение самой мысли. В "Логике" рассматриваются вопросы истинности суждений на основе их взаимосвязи с другими истинными утверждениями. Например, если известно, что "все люди смертны" и "Сократ — человек", то можно заключить, что "Сократ — смертен". В этом примере силлогизма используется дедуктивное правило *modus ponens*. Хотя формальная аксиоматизация логических рассуждений в полном объеме представлена лишь в работах Готлоба Фреге, Бертрана Рассела, Курта Геделя, Алана Тьюринга, Альфреда Тарского и других, корни этих работ можно проследить вплоть до Аристотеля.

Идеи Ренессанса, основанные на греческой традиции, дали толчок развитию иного, мощного представления о человечестве и его роли в природе. На смену мистицизму как средству объяснения вселенной пришел эмпиризм. Часы (а следовательно, и расписание работы фабрик) заменили собой ритм природы для тысяч городских жителей. Большинство современных социальных и физических теорий уходят корнями к идее о возможности математического анализа и постижимости природных или искусственных процессов. В частности, ученые и философы поняли, что мышление само по себе как образ представления знаний является трудным, но принципиальным предметом для научного изучения.

Должно быть, главным событием в развитии современных представлений стала революция, произведенная Коперником, — замена древней геоцентрической модели вселенной, где Земля и другие планеты на самом деле вращаются вокруг Солнца. После столетий господства "очевидности", в которой научное объяснение природы и космоса согласовывалось с религиозным учением и здравым смыслом, была предложена радикально иная (и вовсе не очевидная) модель, объясняющая движение небесных тел. Возможно в первый раз наши *представления о мире рассматривались как фундаментально отличные от их видимости*. Этот разрыв между человеческим разумом и окружающей его реальностью, между понятиями о вещах и самими вещами принципиален для современной теории интеллекта и его организации. Эта брешь была расширена работами Галилея, чьи научные наблюдения еще более расходились с "очевидными" истинами о мире, и чье развитие математики как инструмента для описания мира усилило разрыв между миром и нашими идеями о нем. Именно из этой "бреши" развивалось современное представление о формировании разума: самоанализ стал важным мотивом в литературе, философы начали изучать эпистемологию и математику, и систематизированное применение научного метода стало соперничать с чувствами как орудиями познания мира.

Хотя в XVII и XVIII столетиях было получено немало результатов в эпистемологии и смежных областях, ограничимся рассмотрением работ Рене Декарта. Декарт является центральной фигурой в развитии современных концепций мышления и разума. В своих знаменитых "Размышлениях" Декарт сделал попытку найти основу реальности исключительно методами когнитивной интроспекции. Отвергая информацию, поступающую от органов чувств, как неблагонадежную, Декарт был вынужден подвергнуть сомнению даже существование физического мира и остался наедине с реальностью мысли. Ему пришлось доказывать существование самого себя: "Cogito ergo sum" ("Я мыслю, следовательно, сущест-

вую"). После того как он достоверно установил свое собственное существование как мыслящей сущности, Декарт вывел существование Бога как творца и, в конечном счете, подтвердил реальность физической вселенной как необходимого творения Господа.

Здесь можно сделать два интересных наблюдения. Во-первых, раскол между физическим миром и его интеллектуальным осмыслением стал таким значительным, что появилась возможность рассматривать процесс мышления отдельно от чувственного восприятия или предмета осмысления. Во-вторых, связь между разумом и физическим миром стала столь тонкой, что понадобилось вмешательство всемилостивого Бога, чтобы дать достоверное знание о физическом мире! Это понимание дуализма разума и физического мира пронизывает всю картезианскую мысль, включая открытие аналитической геометрии. Как иначе Декарт мог объединить столь "практичную" область математики, как геометрия, с таким абстрактным математическим основанием, как алгебра?

Почему эта философская дискуссия включена в книгу по искусственному интеллекту? Для ИИ особое значение имеют два важных следствия этих работ.

1. Разделив разум и физический мир, Декарт и его последователи установили, что строение идей о мире не обязательно соответствует изучаемому предмету. На этом основывается методология ИИ, а также эпистемологии, психологии, большей части высшей математики и современной литературы: ментальные процессы существуют сами по себе, подчиняются своим законам и могут изучаться посредством себя же.
2. Поскольку разум и тело оказались разделенными, философы сочли нужным найти способ воссоединить их, ведь взаимодействие между умственным, *res cogitans*, и физическим, *res extensa*, необходимо для человеческого существования.

По поводу проблемы "ума и тела" были написаны миллионы трудов и было предложено множество решений, однако ни одно из них не смогло успешно объяснить очевидные взаимодействия между умственными состояниями и физическими действиями. Наиболее приемлемый ответ на этот вопрос, дающий необходимое основание для изучения ИИ, состоит в том, что ум и тело вовсе не принципиально разные сущности. Согласно этой точке зрения ментальные процессы происходят в таких физических системах, как мозг (или компьютер). Умственные процессы, как и физические, можно, в конечном счете, охарактеризовать с помощью формальной математики. Или, как сказал философ XVII века Гоббс (1651), "мышление есть лишь расчет".

1.1.2. Развитие логики

Поскольку мышление стало рассматриваться как форма вычислений, последующими шагами в его изучении стали формализация и окончательная механизация. В XVIII в. Готфрид Вильгельм фон Лейбниц в работе "Calculus Philosophicus" представил первую систему формальной логики, а также соорудил машину для автоматизации ее вычислений [Leibniz, 1887]. Эйлер в начале восемнадцатого века в своем анализе задачи о кенигсбергских мостах (см. введение в главу 3) создал учение о представлениях, которые абстрактно отражают структуру взаимосвязей реального мира [Euler, 1735].

Формализация теории графов также сделала возможным *поиск в пространстве состояний* (state space search) — основной концептуальный инструмент искусственного интеллекта. Графы можно использовать для моделирования скрытой структуры задачи. Узлы *графа состояний* (state space graph) представляют собой возможные стадии решения задачи; ребра графа отражают умозаключения, ходы в игре или другие шаги в реше-

нии. Решение задачи — это процесс поиска пути к решению на графе состояний (см. раздел 1.3 и главу 3). Описывая все пространство решений задачи, графы состояний предоставляют мощный инструмент для измерения структурированности и сложности проблем, анализа эффективности, корректности и общности стратегий решения.

Как один из основоположников науки исследования операций, а также разработчик первых программируемых механических вычислительных устройств, математик XIX в. Чарльз Бэббидж может также считаться одним из первых практиков искусственного интеллекта [Morrison и Morrison, 1961]. "Разностная машина" Бэббиджа являлась специализированным устройством для вычисления значений некоторых полиномиальных функций и была предшественницей его "аналитической машины". Аналитическая машина, спроектированная, но не построенная при жизни Бэббиджа, была универсальным программируемым вычислительным устройством, которое предвосхитило многие архитектурные положения современных компьютеров.

Описывая аналитическую машину, Ада Лавлейс [Lovelace, 1961], друг Бэббиджа, его помощница и единомышленница, отмечала:

> "Можно сказать, что аналитическая машина плетет алгебраические узоры подобно тому, как станок Жаккарда ткет узоры из цветов и листьев. В этом, как нам кажется, заключается куда больше оригинальности, чем в том, на что могла бы претендовать разностная машина".

Бэббиджа вдохновляло желание применить технологию его времени для освобождения людей от рутины арифметических вычислений. В этом отношении, как и в представлении о вычислительных машинах как механических устройствах, Бэббидж рассуждал всецело с позиций XIX века. Тем не менее его аналитическая машина также основывалась на многих идеях современности, таких как разделение памяти и процессора ("склад" и "мельница", в терминах Бэббиджа), концепция цифровой, а не аналоговой машины и программируемость, основанная на выполнении серий операций, закодированных на картонных перфокартах. Отличительная черта описания Ады Лавлейс и работы Бэббиджа в целом — это отношение к "узорам" алгебраических взаимосвязей как сущностям, которые могут быть изучены, охарактеризованы, наконец, реализованы и подвергнуты механическим манипуляциям без заботы о конкретных значениях, которые проходят через "мельницу" вычислительной машины. Это и есть реализация принципа "абстракции и манипуляции формой", впервые описанного Аристотелем.

Целью создания формального языка для описания мышления задавался также Джордж Буль, математик XIX столетия, чью работу необходимо упомянуть при рассмотрении истоков искусственного интеллекта [Boole, 1847, 1854]. Хотя Буль внес вклад во множество областей математики, его наиболее известным открытием стала математическая формализация законов логики — свершение, сформировавшее самую сердцевину современных компьютерных наук. Роль булевой алгебры в проектировании логический цепей хорошо всем известна, однако цели самого Буля в разработке его системы по духу ближе к современному ИИ. В первой главе книги "Исследование законов мышления, на которых основываются математические теории логики и вероятностей" Буль описывает свои цели следующим образом.

> Исследовать фундаментальные законы таких операций разума, какими совершается рассуждение: дать им выражение в символическом языке исчисления и на этом основании воздвигнуть науку логики и обучать логическому методу; …наконец, из различных элементов истины, усмотренной в этих изысканиях, составить некоторые вероятные догадки касательно природы и склада человеческого ума.

Значимость работы Буля состоит в необычайной силе и простоте предложенной им системы. Три операции: "И" (обозначаемая $*$ или $\wedge$), "ИЛИ" (обозначаемая + или $\vee$) и "НЕ" (обозначаемая символом $\neg$) составляют ядро его логического исчисления. Эти операции стали базой для последующего развития формальной логики, включая разработку современных компьютеров. Сохраняя значения этих символов практически идентичными соответствующим логическим операциям, Буль отмечал, что "символы логики относятся к специальному закону, к которому символы количества как таковые не имеют отношения". Этот "закон" утверждает, что для каждого элемента X алгебры $X*X=X$ (поскольку мы знаем истинность чего-либо, повторение не может изменить это знание). Это привело к ограничению булевых значений всего до двух чисел, которые удовлетворяют этому уравнению, — 1 и 0. Стандартные определения операций булевого умножения (И) и сложения (ИЛИ) следуют из этих соображений.

Булева система не только легла в основу двоичной арифметики, но и показала, что необычайно простая формальная система может передать полную мощь логики. Это предположение и система, разработанная Булем для демонстрации этого факта, стали фундаментом для всех попыток современности формализовать логику, от работы [Whitehead и Russell, 1950], последующих работ Тьюринга и Геделя до современных систем автоматических рассуждений.

Готлоб Фреге (Frege) в своих "Основах арифметики" [Frege, 1884] создал ясный и точный язык спецификации для описания основ арифметики. С помощью этого языка Фреге формализовал многие вопросы, затронутые ранее в аристотелевской "Логике". Язык Фреге, сейчас именуемый *исчислением предикатов первого порядка*, служит инструментом для записи теорем и задания значений истинности, которые образуют элементы математических умозаключений и описывают аксиоматический базис "смысла" этих выражений. Предполагалось, что формальная система исчисления предикатов, которая включает символы предикатов, теорию функций и квантированных переменных, станет языком для описания математики и ее философских основ. Она также сыграла принципиальную роль в создании теории представления для искусственного интеллекта (см. главу 2). Исчисление предикатов первого порядка обеспечивает средства автоматизации рассуждений: язык для построения выражений, теорию, позволяющую судить об их смысле, и логически безупречное исчисление для вывода новых истинных выражений.

Работа Рассела и Уайтхеда особенно важна для фундаментальных принципов ИИ, поскольку заявленной ими целью было вывести из набора аксиом путем формальных операций всю математику. Хотя многие математические системы строились на основе аксиом, интересно отношение Рассела и Уайтхеда к математике как к чисто формальной системе. Это означает, что аксиомы и теоремы должны рассматриваться исключительно как наборы символов: доказательства должны выводиться лишь посредством применения строго определенных правил для манипулирования такими строками. При этом исключается использование интуиции или "смысла" теорем в качестве основы доказательств. Каждый шаг доказательства следует из строгого применения формальных (синтаксических) правил к аксиомам или уже выведенным теоремам, даже если в традиционных доказательствах этот шаг назывался "очевидным". Смысл, содержащийся в теоремах и аксиомах системы, имеет отношение только к внешнему миру и совершенно не зависит от логического вывода. Такой полностью формальный (реализуемый техническими средствами) подход к математическим умозаключениям предоставил существенную основу для его автоматизации в реальных вычислительных машинах. Логический синтаксис и формальные правила вывода, разработанные Расселом и Уайтхедом,

лежат в основе систем автоматического доказательства теорем, рассматриваемых в главе 12, а также составляют теоретические основы искусственного интеллекта.

Альфред Тарский (Tarski) — еще один математик, чьи работы сыграли принципиальную роль в процессе формирования искусственного интеллекта. Тарский [Tarski, 1944, 1956] создал *теорию ссылок* (theory of reference), согласно которой *правильно построенные формулы* (well-formed formulae) Фреге или Рассела–Уайтхеда определенным образом ссылаются на объекты реального мира (см. главу 2). Эта концепция лежит в основе большинства теорий формальной семантики. В работе "Семантическая концепция истинности и основание семантики" Тарский описывает свою теорию ссылок и взаимосвязей между значениями истинности. Современные исследователи компьютерных наук связали эту теорию с языками программирования и другими компьютерными реалиями [Burstall и Darlington, 1977].

Хотя в XVIII–XIX вв. и начале XX в. формализация науки и математики создала интеллектуальные предпосылки для изучения искусственного интеллекта, он не стал жизнеспособной научной дисциплиной до появления цифровых вычислительных машин. К концу 1940-х гг. электронные цифровые компьютеры продемонстрировали свои возможности в предоставлении памяти и процессорной мощности, требуемой для интеллектуальных программ. Стало возможным реализовать формальные системы рассуждений в машине и эмпирически испытать их достаточность для проявления разумности. Существенной составляющей теории искусственного интеллекта является взгляд на цифровые компьютеры как на средство создания и проверки теорий интеллекта.

Но цифровые компьютеры — не только рабочая лошадка для испытания теорий интеллекта. Их архитектура наталкивает на специфичное представление таких теорий: интеллект — это способ обработки информации. Например, концепция поиска как методики решения задач обязана своим появлением в бо́льшей степени последовательному характеру компьютерных операций, нежели какой-либо биологической модели интеллекта. Большинство программ ИИ представляют знания на некотором формальном языке, а затем обрабатывают их в соответствии с алгоритмами, следуя заложенному еще фон Нейманом принципу разделения данных и программы. Формальная логика возникла как важный инструмент представления для исследований ИИ, равно как теория графов играет неоценимую роль в анализе пространства, а также предоставляет основу для семантических сетей и схожих моделей. Эти методы и формализмы детально обсуждаются в последующих главах книги. Здесь они упоминаются для подчеркивания симбиотических отношений между цифровыми компьютерами и теоретическими основами искусственного интеллекта.

Мы часто забываем, что инструменты, которые мы создаем для своих целей, влияют своим устройством и ограничениями на формирование наших представлений о мире. Такое казалось бы стесняющее наш кругозор взаимодействие является важным аспектом развития человеческого знания: инструмент (а научные теории, в конечном счете, тоже инструменты) создается для решения конкретной проблемы. По мере применения и совершенствования инструмент подсказывает другие способы его использования, которые приводят к новым вопросам и, в конце концов, разработке новых инструментов.

1.1.3. Тест Тьюринга

Одна из первых работ, посвященных вопросу о машинном разуме в отношении современных цифровых компьютеров, "Вычислительные машины и интеллект" была написана в 1950 г. британским математиком Аланом Тьюрингом и опубликована в журнале

“Mind” [Turing, 1950]. Она не теряет актуальности, как по части аргументов против возможности создания разумной вычислительной машины, так и по части ответов на них. Тьюринг, известный в основном благодаря своим трудам по теории вычислимости, рассмотрел вопрос о том, можно ли заставить машину действительно думать. Отмечая, что фундаментальная неопределенность в самом вопросе (что такое “думать”? что такое “машина”?) исключает возможность рационального ответа, он предложил заменить вопрос об интеллекте более четко определенным эмпирическим тестом.

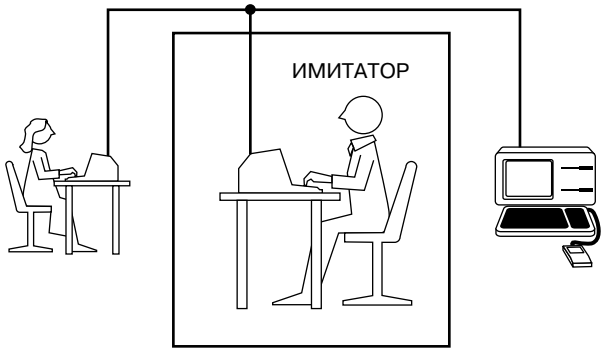

Рис. 1.1. Тест Тьюринга

Тест Тьюринга сравнивает способности предположительно разумной машины со способностями человека — лучшим и единственным стандартом разумного поведения. В тесте, который Тьюринг назвал “имитационной игрой”, машину и ее человеческого соперника (следователя) помещают в разные комнаты, отделенные от комнаты, в которой находится “имитатор” (рис. 1.1). Следователь не должен видеть их или говорить с ними напрямую — он сообщается с ними исключительно с помощью текстового устройства, например, компьютерного терминала. Следователь должен отличить компьютер от человека исключительно на основе их ответов на вопросы, задаваемые через это устройство. Если же следователь не может отличить машину от человека, тогда, утверждает Тьюринг, машину можно считать разумной.

Изолируя следователя от машины и другого человека, тест исключает предвзятое отношение — на решение следователя не будет влиять вид машины или ее электронный голос. Следователь волен задавать любые вопросы, не важно, насколько окольные или косвенные, пытаясь раскрыть “личность” компьютера. Например, следователь может попросить обоих подопытных осуществить довольно сложный арифметический подсчет, предполагая, что компьютер скорее даст верный ответ, чем человек. Чтобы обмануть эту стратегию, компьютер должен знать, когда ему следует выдать ошибочное число, чтобы показаться человеком. Чтобы обнаружить человеческое поведение на основе эмоциональной природы, следователь может попросить обоих субъектов высказаться по поводу стихотворения или картины. Компьютер в таком случае должен знать об эмоциональном складе человеческих существ.

Этот тест имеет следующие важные особенности.

1. Дает объективное понятие об интеллекте, т.е. реакции заведомо разумного существа на определенный набор вопросов. Таким образом, вводится стандарт для определения интеллекта, который предотвращает неминуемые дебаты об “истинности” его природы.

2. Препятствует заведению нас в тупик сбивающими с толку и пока безответными вопросами, такими как: должен ли компьютер использовать какие-то конкретные внутренние процессы, или же должна ли машина по-настоящему осознавать свои действия.
3. Исключает предвзятость в пользу живых существ, заставляя опрашивающего сфокусироваться исключительно на содержании ответов на вопросы.

Благодаря этим преимуществам, тест Тьюринга представляет собой хорошую основу для многих схем, которые используются на практике для испытания современных интеллектуальных программ. Программа, потенциально достигшая разумности в какой-либо предметной области, может быть испытана сравнением ее способностей по решению данного множества проблем со способностями человеческого эксперта. Этот метод испытания всего лишь вариация на тему теста Тьюринга: группу людей просят сравнить "вслепую" ответы компьютера и человека. Как видим, эта методика стала неотъемлемым инструментом как при разработке, так и при проверке современных экспертных систем.

Тест Тьюринга, несмотря на свою интуитивную притягательность, уязвим для многих оправданных нападок. Одно из наиболее слабых мест — пристрастие в пользу чисто символьных задач. Тест не затрагивает способностей, требующих навыков перцепции или ловкости рук, хотя подобные аспекты являются важными составляющими человеческого интеллекта. Иногда же, напротив, тест Тьюринга обвиняют в попытках втиснуть машинный интеллект в форму интеллекта человеческого. Быть может, машинный интеллект просто настолько отличается от человеческого, что проверять его человеческими критериями — фундаментальная ошибка? Нужна ли нам, в самом деле, машина, которая бы решала математические задачи так же медленно и неточно, как человек? Не должна ли разумная машина извлекать выгоду из своих преимуществ, таких как большая, быстрая, надежная память, и не пытаться сымитировать человеческое познание? На самом деле, многие современные практики ИИ (например [Ford и Hayes, 1995]) говорят, что разработка систем, которые бы выдерживали всесторонний тест Тьюринга, — это ошибка, отвлекающая нас от более важных, насущных задач: разработки универсальных теорий, объясняющих механизмы интеллекта людей и машин и применение этих теорий к проектированию инструментов для решения конкретных практических проблем. Все же тест Тьюринга представляется нам важной составляющей в тестировании и "аттестации" современных интеллектуальных программ.

Тьюринг также затронул проблему осуществимости построения интеллектуальной программы на базе цифрового компьютера. Размышляя в терминах конкретной вычислительной модели (электронной цифровой машины с дискретными состояниями), он сделал несколько хорошо обоснованных предположений касательно ее объема памяти, сложности программы и основных принципов проектирования такой системы. Наконец, он рассмотрел множество моральных, философских и научных возражений возможности создания такой программы средствами современной технологии. Отсылаем читателя к статье Тьюринга за познавательным и все еще актуальным изложением сути споров о возможностях интеллектуальных машин.

Два возражения, приведенных Тьюрингом, стоит рассмотреть детально. "Возражение леди Лавлейс", впервые сформулированное Адой Лавлейс, сводится к тому, что компьютеры могут делать лишь то, что им укажут, и, следовательно, не могут выполнять оригинальные (читай: разумные) действия. Однако экспертные системы (см. подраздел 1.2.3 и главу 7), особенно в области диагностики, могут формулировать выводы, которые не были заложены в них разработчиками. Многие исследователи считают, что творческие способности можно реализовать программно.

Другое возражение, “аргумент естественности поведения”, связано с невозможностью создания набора правил, которые бы говорили индивидууму, что в точности нужно делать при каждом возможном стечении обстоятельств. Действительно, гибкость, позволяющая биологическому разуму реагировать практически на бесконечное количество различных ситуаций приемлемым, если даже и не оптимальным образом — отличительная черта разумного поведения. Справедливо замечание, что управляющая логика, используемая в большинстве традиционных компьютерных программ, не проявляет великой гибкости или силы воображения, но неверно, что все программы должны писаться подобным образом. Бо́льшая часть работ в сфере ИИ за последние 25 лет была направлена на разработку таких языков программирования и моделей, призванных устранить упомянутый недостаток, как продукционные системы, объектные системы, сетевые представления и другие модели, обсуждаемые в этой книге.

Современные программы ИИ обычно состоят из набора модульных компонентов, или правил поведения, которые не выполняются в жестко заданном порядке, а активизируются по мере надобности в зависимости от структуры конкретной задачи. Системы обнаружения совпадений позволяют применять общие правила к целому диапазону задач. Эти системы необычайно гибки, что позволяет относительно маленьким программам проявлять разнообразное поведение в широких пределах, реагируя на различные задачи и ситуации.

Можно ли довести гибкость таких программ до уровня живых организмов, все еще предмет жарких споров. Нобелевский лауреат Герберт Саймон сказал, что большей частью своеобразие и изменчивость поведения, присущие живым существам, возникли скорее благодаря сложности их окружающей среды, чем благодаря сложности их внутренних “программ”. В [Simon, 1981] Саймон описывает муравья, петляющего по неровной, пересеченной поверхности. Хотя путь муравья кажется довольно сложным, Саймон утверждает, что цель муравья очень проста: вернуться как можно скорее в колонию. Изгибы и повороты его пути вызваны встречаемыми препятствиями. Саймон заключает, что:

> “Муравей, рассматриваемый в качестве проявляющей разумное поведение системы, на самом деле очень прост. Кажущаяся сложность его поведения в большей степени отражает сложность среды, в которой он существует”.

Эта идея, если удастся доказать применимость ее к организмам с более сложным интеллектом, составит сильный аргумент в пользу простоты, а следовательно, постижимости интеллектуальных систем. Любопытно, что, применив эту идею к человеку, мы придем к выводу об огромной значимости культуры в формировании интеллекта. Интеллект, похоже, не взращивается во тьме, как грибы. Для его развития необходимо взаимодействие с достаточно богатой окружающей средой. Культура так же необходима для создания человеческих существ, как и человеческие существа для создания культуры. Эта мысль не умаляет могущества наших интеллектов, но подчеркивает удивительное богатство и связь различных культур, сформировавших жизни отдельных людей. Фактически на идее о том, что интеллект возникает из взаимодействий индивидуальных элементов общества, основывается подход к ИИ, представленный в следующем разделе.

1.1.4. Биологические и социальные модели интеллекта: агенты

Итак, мы рассмотрели математический подход к задаче построения интеллектуальных устройств, подразумевающий, что основой самого интеллекта являются логические умозаключения, а также основанный на “объективности” самих логических рассуждений. Этот взгляд на знание, язык и мышление отражает традицию рационализма западной философии,

развитую в работах Платона, Галилея, Декарта, Лейбница и многих других философов, упомянутых ранее в этой главе. Также он отражает неявные предположения теста Тьюринга, особенно его взгляд на символьные рассуждения как критерий интеллекта, и веру, что "лобовое" сравнение с человеческим поведением пригодно для подтверждения интеллекта машины.

Опора на логику как способ представления языка и логические выводы как основной механизм разумных рассуждений настолько доминирует в западной философии, что их "истинность" часто кажется очевидной и неоспоримой. Поэтому не удивительно, что подходы, основанные на этих предположениях, главенствуют в науке искусственного интеллекта от ее зарождения до сегодняшнего дня.

Во второй половине XX века устои рационализма пошатнулись. Философский релятивизм в разных своих формах задавался вопросом об объективном базисе языка, науки, общества и самой мысли. Философия поздних работ Виттгенштейна [Wittgenstein, 1953] вынудила пересмотреть понятие смысла в естественных и формальных языках. Труды Геделя и Тьюринга подвергли сомнению основания самой математики. Постмодернистские идеи изменили наши взгляды на значимость и ценность в художественном и социальном контекстах. Искусственный интеллект также стал жертвой подобной критики. Действительно, трудности, которые встали на пути ИИ к его целям, часто рассматриваются как свидетельства ошибочности рационалистического взгляда [Winograd и Flores, 1986], [Lakoff и Johnson, 1999].

Две философские традиции — Виттгенштейна с Хассерлом [Husserl, 1970, 1972] и Хайдеггера [Heidegger, 1962] являются основополагающими в этом пересмотре западной философии. В своей работе Виттгенштейн затронул многие допущения рационалистской традиции, включая основания языка, науки и знания. Естественный язык был главным предметом анализа Виттгенштейна. Этот философ опровергает мнение, что смысл человеческого языка можно вывести из каких-либо объективных основ.

В трудах Виттгенштейна, как и в теории речи (speech act theory), развитой Остином [Austin, 1962] и его последователями [Grice, 1975], [Searle, 1969], значение любого высказывания зависит от человеческого, культурного контекста. Значение слова "сиденье", к примеру, зависит от наличия физического объекта, который можно применить для сидения на нем, а также культурных соглашений об использовании сидений. Когда, например, большой плоский камень можно назвать сиденьем? Почему нелепо так называть королевский трон? Какая разница между человеческим пониманием "сиденья" и пониманием кота или собаки, которые в человеческом смысле сидеть не могут? Атакуя основы смысла, Виттгенштейн утверждал, что мы должны рассматривать использование языка посредством выбора и действий в изменчивом культурном контексте. Виттгенштейн даже распространил свою критику на науку и математику, утверждая, что они в такой же мере общественные конструкции, как и языки.

Хассерл, отец феноменологии, рассматривал абстракции как объекты, укоренившиеся в конкретном "жизненном мире": рационалистская модель отодвигает конкретный поддерживающий ее мир на второй план. Для Хассерла, как и для его ученика Хайдеггера и их сторонника Мерло-Понти [Merleau-Ponty, 1962], интеллект заключался не в знании истины, а в знании, как вести себя в постоянно меняющемся и развивающемся мире. Таким образом, в экзистенциалистско-феноменологической традиции интеллект рассматривается скорее с точки зрения выживания в мире, чем как набор логических утверждений о мире (в сочетании со схемой вывода).

Многие авторы, например Дрейфусы [Dreyfus и Dreyfus, 1985], а также Виноград и Флорес [Winograd и Flores, 1986], опирались на работы Виттгенштейна, Хассерла и Хай-

деггера в своей критике ИИ. Хотя многие практики ИИ продолжают разработку рационально-логической программной системы (также известной как GOFAI, или Good Old Fashioned AI — старый-добрый ИИ), все возрастающее число исследователей этой области, приняв во внимание эту критику, строят новые занимательные модели интеллекта. Придерживаясь идей Витгенштейна об антропологических и культурных корнях знания, они обратились к социальным моделям интеллектуального поведения, иногда называемым *ситуативными*.

Пример альтернативы логическому подходу — исследования в области коннекционистского обучения (см. подраздел 1.2.9 и главу 10), в которых логике и работе рационального разума уделяется мало внимания, но сделана попытка достичь разумности посредством моделирования архитектуры реального мозга. В нейронных моделях интеллекта упор делается на способность мозга адаптироваться к миру, в котором он существует, с помощью изменений связей между отдельными нейронами. Знание в таких системах не выражается явными логическими конструкциями, а представляется в неявной форме, как свойство конфигураций таких взаимосвязей.

Иная модель интеллекта, заимствованная из биологии, навеяна процессами адаптации видов к окружающей среде. В разработках искусственной жизни и генетических алгоритмов (см. главу 11) принципы биологической эволюции применяются для решения сложных проблем. Такие программы не решают задачи посредством логических рассуждений. Они порождают популяции соревнующихся между собой решений-кандидатов и заставляют их совершенствоваться с помощью процессов, имитирующих биологическую эволюцию: неудачные кандидаты на решения отмирают, в то время как подающие надежды выживают и воспроизводятся путем создания новых решений из частей “успешных” родителей.

Социальные системы дают еще одно модельное представление интеллекта с помощью глобального поведения, которое позволяет им решать проблемы, которые бы не удалось решить отдельным их членам. Например, хотя ни один индивидуум не в состоянии точно предсказать количество буханок хлеба, которое потребит в заданный день Нью-Йорк, система всех нью-йоркских пекарен отлично справляется со снабжением города хлебом и делает это с минимальными затратами. Рынок акций отлично устанавливает относительную ценность сотен компаний, хотя каждый отдельный инвестор имеет лишь ограниченное знание о нескольких компаниях. Можно привести также пример из современной науки. Отдельные исследователи из университетской, производственной или правительственной среды сосредоточиваются на решении общих проблем. С помощью конференций и журналов, служащих основным средством сообщения, важные для общества в целом проблемы рассматриваются и решаются отдельными агентами, работающими отчасти независимо, хотя прогресс во многих случаях также направляется субсидиями.

Эти примеры имеют два общих аспекта. Во-первых, корни интеллекта связаны с культурой и обществом, а следовательно, разум является *эмерджентным* (emergent). Во-вторых, разумное поведение формируется совместными действиями большого числа очень простых взаимодействующих полуавтономных индивидуумов, или агентов. Являются агенты нервными клетками, индивидуальными особями биологического вида или же отдельными личностями в обществе, их взаимодействие создает интеллект.

Рассмотрим основные аспекты агентских и эмерджентных взглядов на интеллект.

1. Агенты автономны или полуавтономны. Следовательно, у каждого агента есть определенный круг подзадач, причем он располагает малым знанием (или вовсе не располагает знанием) о том, что делают другие агенты или как они это делают. Каждый агент выполняет свою независимую часть решения проблемы и либо

выдает собственно результат (что-то совершает) либо сообщает результат другим агентам.

2. Агенты являются “внедренными”. Каждый агент чувствителен к своей окружающей среде и (обычно) не знает о состоянии полной области существования агентов. Таким образом, знание агента ограничено его текущими задачами: “файл, который я обрабатываю” или “стенка рядом со мной”, агент не владеет информацией обо всех файлах одновременно или физических границах предметной области.
3. Агенты взаимодействуют. Они формируют коллектив индивидуумов, которые сотрудничают над решением задачи. В этом смысле их можно рассматривать как “сообщество”. Как и в человеческом обществе, знания, умения и обязанности распределяются среди отдельных индивидуумов.
4. Сообщество агентов структурировано. В большинстве агентно-ориентированных методов решения проблем каждый индивидуум, работая со своим собственным окружением и навыками, координирует общий ход решения с другими агентами. Таким образом, окончательное решение можно назвать не только коллективным, но и кооперативным.
5. Наконец, явление интеллекта в этой среде является “эмерджентным”. Хотя индивидуальные агенты обладают некоторыми совокупностями навыков и обязанностей, общий, совместный, результат сообщества агентов следует рассматривать как нечто большее, чем сумма отдельных вкладов. Интеллект рассматривается как явление, возникающее в сообществе, а не как свойство отдельного агента.

Основываясь на этих наблюдениях, определим агента как элемент сообщества, который может воспринимать (часто ограниченно) аспекты своего окружения и взаимодействовать с этой окружающей средой либо непосредственно, либо путем сотрудничества с другими агентами. Большинство интеллектуальных методов решений требуют наличия разнообразных агентов. Это могут быть простые агенты-механизмы, задача которых — собирать и передавать информацию; агенты-координаторы, которые обеспечивают взаимодействие между другими агентами; агенты поиска, которые перебирают пакеты информации и возвращают какие-то избранные частицы; обучающие агенты, которые на основе полученной информации формируют обобщающие концепции; и агенты, принимающие решения, которые раздают задания и делают выводы на основе ограниченной информации и обработки. Возвращаясь к старому определению интеллекта, агенты можно рассматривать как механизмы, обеспечивающие выработку решения в условиях ограниченных ресурсов и процессорных мощностей.

Для разработки и построения таких сообществ необходимы следующие компоненты.

1. Структуры для представления информации.
2. Стратегии поиска в пространстве альтернативных решений.
3. Архитектура, обеспечивающая взаимодействие агентов.

В последующих главах, в частности в разделе 6.4, приводятся рекомендации для создания средств поддержки таких агентских сообществ.

Это предварительное рассмотрение возможностей теории автоматизированного интеллекта ни в коей мере не ставит себе целью преувеличить достижения, полученные до настоящего времени или преуменьшить работу, которую еще предстоит проделать. В этой книге постоянно подчеркивается, как важно четко представлять себе границы на-

ших текущих возможностей и сегодняшних свершений. Например, очень ограниченные успехи сделаны в построении программ, которые могут "учиться" в интересном для человека смысле. Также очень скромны достижения в моделировании семантической сложности естественных языков, к примеру, английского. Даже в фундаментальных вопросах, таких как организация знаний или управление сложностью и корректностью большой компьютерной программы (например, большой базы знаний), требуются значительные дальнейшие исследования. Основанные на знаниях системы, хотя и достигли коммерческого успеха, имеют все еще много ограничений в качестве и общности своих выводов. В частности, они не способны рассуждать на основе "здравого смысла" (commonsense reasoning) или же проявлять знания о простейших свойствах реального мира, например, о том, как изменяются со временем разные вещи.

Но следует сохранять трезвый взгляд на вещи. Обозреть достижения искусственного интеллекта легче, честно представляя предстоящую работу. В следующем разделе мы рассмотрим некоторые сферы исследований и разработок ИИ.

1.2. Обзор прикладных областей искусственного интеллекта

Аналитическая машина не претендует на создание чего-либо нового. Ее способности не превосходят наших знаний о том, как приказать ей что-либо исполнить…

— Ада Байрон (Ada Byron), графиня Лавлейс

Прости, Дейв, я не могу позволить тебе это сделать…

— HAL 9000, компьютер из фильма *2001: Космическая одиссея*

Вернемся к заявленной цели дать определение искусственному интеллекту путем обозрения стремлений и достижений исследователей этой области. Две наиболее фундаментальные проблемы, занимающие разработчиков ИИ, — это *представление знаний* (knowledge representation) и *поиск* (search). Первая относится к проблеме получения новых знаний с помощью формального языка, подходящего для компьютерных манипуляций, всего спектра знаний, требуемых для формирования разумного поведения. В главе 2 рассматривается исчисление предикатов как язык описания свойств и отношений между объектами предметной области. Здесь для решения нужны скорее рассуждения качественного характера, чем арифметические расчеты. В главах 6–8 обсуждаются языки, разработанные для отражения неопределенностей и структурной сложности таких областей, как рассуждения на основе "здравого смысла" и понимание естественных языков. В главах 14 и 15 демонстрируется использование языков LISP и PROLOG для реализаций этих представлений.

Поиск — это метод решения проблемы, в котором систематически просматривается пространство *состояний задачи* (problem states), т.е. альтернативных стадий ее решения. Примеры состояний задачи: различные размещения фигур на доске в игре или же промежуточные шаги логического обоснования. Затем в этом пространстве альтернативных решений производится перебор в поисках окончательного ответа. Ньюэлл и Саймон [Newell и Simon, 1976] утверждают, что эта техника лежит в основе человеческого способа решения различных задач. Действительно, когда игрок в шахматы анализирует последствия различных ходов или врач обдумывает различные альтернативные диагнозы,

они производят перебор среди альтернатив. Результаты применения этой модели и средства ее реализации обсуждаются в главах 3, 4, 5 и 16.

Как и бо́льшая часть наук, ИИ разбивается на множество поддисциплин, которые, разделяя основной подход к решению проблем, нашли себе различные применения. Очертим в этом разделе некоторые из основных сфер применения этих отраслей и их вклад в искусственный интеллект вообще.

1.2.1. Ведение игр

Многие ранние исследования в области поиска в пространстве состояний совершались на основе таких распространенных настольных игр, как шашки, шахматы и пятнашки. Вдобавок к свойственному им "интеллектуальному" характеру такие игры имеют некоторые свойства, делающие их идеальным объектом для экспериментов. Большинство игр ведутся с использованием четко определенного набора правил: это позволяет легко строить пространство поиска и избавляет исследователя от неясности и путаницы, присущих менее структурированным проблемам. Позиции фигур легко представимы в компьютерной программе, они не требуют создания сложных формализмов, необходимых для передачи семантических тонкостей более сложных предметных областей. Тестирование игровых программ не порождает никаких финансовых или этических проблем. Поиск в пространстве состояний — принцип, лежащий в основе большинства исследований в области ведения игр, — представлен в главах 3 и 4.

Игры могут порождать необычайно большие пространства состояний. Для поиска в них требуются мощные методики, определяющие, какие альтернативы следует рассматривать. Такие методики называются *эвристиками* и составляют значительную область исследований ИИ. Эвристика — стратегия полезная, но потенциально способная упустить правильное решение. Примером эвристики может быть рекомендация проверять, включен ли прибор в розетку, прежде чем делать предположения о его поломке, или выполнять рокировку в шахматной игре, чтобы попытаться уберечь короля от шаха. Бо́льшая часть того, что мы называем разумностью, по-видимому, опирается на эвристики, которые люди используют в решении задач.

Поскольку у большинства из нас есть опыт в этих простых играх, можно попробовать разработать свои эвристики и испытать их эффективность. Для этого нам не нужны консультации экспертов в каких-то темных для непосвященных областях, вроде медицины или математики. Поэтому игры являются хорошей основой для изучения эвристического поиска. Глава 4 рассказывает об эвристиках на примере этих простых игр; в главе 7 их использование распространяется на построение экспертных систем. Программы ведения игр, несмотря на их простоту, ставят перед исследователями новые вопросы, включая вариант, при котором ходы противника невозможно детерминировано предугадать (см. главу 8). Наличие противника усложняет структуру программы, добавляя в нее элемент непредсказуемости и потребность уделять внимание психологическим и тактическим факторам игровой стратегии.

1.2.2. Автоматические рассуждения и доказательство теорем

Можно сказать, что автоматическое доказательство теорем — одна из старейших частей искусственного интеллекта, корни которой уходят к системам Logic Theorist (логический теоретик) Ньюэлла и Саймона [Newell и Simon, 1963*a*] и General Problem

Solver (универсальный решатель задач) [Newell и Simon, 1963*б*] и далее, к попыткам Рассела и Уайтхеда построить всю математику на основе формальных выводов теорем из начальных аксиом. В любом случае эта ветвь принесла наиболее богатые плоды. Благодаря исследованиям в области доказательства теорем были формализованы алгоритмы поиска и разработаны языки формальных представлений, такие как исчисление предикатов (см. главу 2) и логический язык программирования PROLOG (глава 14).

Привлекательность автоматического доказательства теорем основана на строгости и общности логики. В формальной системе логика располагает к автоматизации. Разнообразные проблемы можно попытаться решить, представив описание задачи и существенную относящуюся к ней информацию в виде логических аксиом и рассматривая различные случаи задачи как теоремы, которые нужно доказать. Этот принцип лежит в основе автоматического доказательства теорем и систем математических обоснований (см. главу 12).

К сожалению, в ранних пробах написать программу для автоматического доказательства не удалось разработать систему, которая бы единообразно решала сложные задачи. Это было обусловлено способностью любой относительно сложной логической системы сгенерировать бесконечное количество доказуемых теорем: без мощных методик (эвристик), которые бы направляли поиск, программы доказывали большие количества не относящихся к делу теорем, пока не натыкались на нужную. Из-за этой неэффективности многие утверждают, что чисто формальные синтаксические методы управления поиском в принципе не способны справиться с такими большими пространствами, и единственная альтернатива этому — положиться на неформальные, специально подобранные к случаю (лат. “ad hoc”) стратегии, как это, похоже, делают люди. Это один из подходов, лежащих в основе экспертных систем (см. главу 7), и он оказался достаточно плодотворным.

Все же привлекательность рассуждений, основанных на формальной логике, слишком сильна, чтобы ее игнорировать. Многие важные проблемы, такие как проектирование и проверка логических цепей, проверка корректности компьютерных программ и управление сложными системами, по-видимому, поддаются такому подходу. Вдобавок исследователям автоматического доказательства удалось разработать мощные эвристики, основанные на оценке синтаксической формы логического выражения, которые в результате понижают сложность пространства поиска, не прибегая к используемым людьми методам “ad hoc”.

Еще одной причиной неувядающего интереса к автоматическому доказательству теорем является понимание, что системе не обязательно решать особо сложные проблемы без человеческого вмешательства. Многие современные программы доказательств работают как умные помощники, предоставляя людям разбивать задачи на подзадачи и продумывать эвристики для перебора в пространстве возможных обоснований. Программа для автоматического доказательства затем решает более простые задачи доказательства лемм, проверки менее существенных предположений и дополняет формальные аспекты доказательства, очерченного человеком [Boyer и Moore, 1979], [Bundy, 1988], [Veroff, 1997].

1.2.3. Экспертные системы

Одним из главных достижений ранних исследований по ИИ стало осознание важности специфичного для предметной области (domain-specific) знания. Врач, к примеру, хорошо диагностирует болезни не потому, что он располагает некими врожденными общими способностями к решению задач, а потому, что многое знает о медицине. Точно так же геолог эффективно находит залежей ископаемых, потому что он способен применить богатые теоретические и практические знания о геологии к текущей проблеме. Экспертное знание — это

сочетание теоретического понимания проблемы и набора эвристических правил для ее решения, которые, как показывает опыт, эффективны в данной предметной области. Экспертные системы создаются с помощью заимствования знаний у человеческого эксперта и кодирования их в форму, которую компьютер может применить к аналогичным проблемам.

Стратегии экспертных систем основаны на знаниях человека-эксперта. Хотя многие программы пишутся самими носителями знаний о предметной области, большинство экспертных систем являются плодом сотрудничества между таким экспертом, как врач, химик, геолог или инженер, и независимым специалистом по ИИ. Эксперт предоставляет необходимые знания о предметной области, описывая свои методы принятия решений и демонстрируя эти навыки на тщательно отобранных примерах. Специалист по ИИ, или *инженер по знаниям* (knowledge engineer), как часто называют разработчиков экспертных систем, отвечает за реализацию этого знания в программе, которая должна работать эффективно и внешне разумно. Экспертные способности программы проверяют, давая ей решать пробные задачи. Эксперт подвергает критике поведение программы, и в ее базу знаний вносятся необходимые изменения. Процесс повторяется, пока программа не достигнет требуемого уровня работоспособности.

Одной из первых систем, использовавших специфичные для предметной области знания, была система DENDRAL, разработанная в Стэнфорде в конце 1960-х [Lindsay и др., 1980]. DENDRAL была задумана для определения строения органических молекул из химических формул и спектрографических данных о химических связях в молекулах. Поскольку органические молекулы обычно очень велики, число возможных структур этих молекул также весьма внушительно. DENDRAL решает проблему большого пространства перебора, применяя эвристические знания экспертов-химиков к решению задачи определения структуры. Методы DENDRAL оказались весьма работоспособными. Она методично находит правильное строение из миллионов возможных всего за несколько попыток. Данный подход оказался столь эффективным, что “потомки” этой системы до сих пор используются в химических и фармацевтических лабораториях по всему миру.

Программа DENDRAL одной из первых использовала специфичное знание для достижения уровня эксперта в решении задач, однако методика современных экспертных систем связана с другой программой — MYCIN [Buchanan и Shortliffe, 1984]. В ней использовались знания экспертов медицины для диагностики и лечения спинального менингита и бактериальных инфекций крови.

Программа MYCIN, разработанная в Стэнфорде в середине 1970-х, одной из первых обратилась к проблеме принятия решений на основе ненадежной или недостаточной информации. Она выводит ясные и логичные пояснения своих рассуждений, используя структуру управляющей логики, соответствующую специфике предметной области, и критерии для надежной оценки своей работы. Многие методики разработки экспертных систем, использующиеся сегодня, были впервые разработаны в рамках проекта MYCIN (см. главу 7).

К числу других классических экспертных систем относится программа PROSPECTOR, определяющая предполагаемые рудные месторождения и их типы, основываясь на геологических данных о местности [Duda и др., 1979*а*, 1979*б*]; программа INTERNIST, применяемая для диагностики в сфере медицины внутренних органов; программа Dipmeter Advisor, интерпретирующая протоколы бурения нефтяных скважин [Smith и Baker, 1983]; и XCON, используемая для настройки компьютеров VAX. Программа XCON была разработана в 1981 г., и одно время все машины VAX, распространяемые компанией Digital Equipment, настраивались этой программой. Многочисленные экспертные системы решают в настоящее время задачи в таких областях, как медицина, образование, бизнес, дизайн и научные исследования [Waterman, 1986], [Durkin, 1994].

Интересно отметить, что большинство экспертных систем были написаны для специализированных предметных областей. Эти области довольно хорошо изучены и располагают четко определенными стратегиями принятия решений. Проблемы, определенные на нечеткой основе "здравого смысла", подобными средствами решить сложнее. Несмотря на воодушевляющие перспективы экспертных систем, было бы ошибкой переоценивать возможности этой технологии. Основные проблемы перечислены ниже.

1. Трудности в передаче "глубоких" знаний предметной области. Системе MYCIN, к примеру, не достает действительного знания человеческой физиологии. Она не знает, какова функция кровеносной системы или спинного мозга. Ходит предание, что однажды, подбирая лекарство для лечения менингита, MYCIN спросила, беремен ли пациент, хотя ей указали, что пациент мужского пола. Было это на самом деле или нет, неизвестно, но это хорошая иллюстрация потенциальной ограниченности знаний экспертной системы.
2. Недостаток здравомыслия и гибкости. Если людей поставить перед задачей, которую они не в состоянии решить немедленно, то они обычно исследуют сперва основные принципы и вырабатывают какую-то стратегию для подхода к проблеме. Экспертным системам этой способности не хватает.
3. Неспособность предоставлять осмысленные объяснения. Поскольку экспертные системы не владеют глубоким знанием своей предметной области, их пояснения обычно ограничиваются описанием шагов, которые система предприняла в поиске решения. Но они зачастую не могут пояснить, "почему" был выбран конкретный подход.
4. Трудности в тестировании. Хотя обоснование корректности любой большой компьютерной системы достаточно трудоемко, экспертные системы проверять особенно тяжело. Это серьезная проблема, поскольку технологии экспертных систем применяются для таких критичных задач, как управление воздушным движением, ядерными реакторами и системами оружия.
5. Ограниченные возможности обучения на опыте. Сегодняшние экспертные системы делаются "вручную"; производительность разработанной системы не будет возрастать до следующего вмешательства программистов. Это заставляет серьезно усомниться в разумности таких систем.

Несмотря на эти ограничения, экспертные системы доказали свою ценность во многих важных приложениях. Будем надеяться, что недоработки сподвигнут студентов заняться этой важной отраслью компьютерных наук. Экспертные системы — одна из основных тем этой книги. Они подробно обсуждаются в главах 6 и 7.

1.2.4. Понимание естественных языков и семантическое моделирование

Одной из долгосрочных целей искусственного интеллекта является создание программ, способных понимать человеческий язык и строить фразы на нем. Способность применять и понимать естественный язык является фундаментальным аспектом человеческого интеллекта, а его успешная автоматизация привела бы к неизмеримой эффективности самих компьютеров. Многие усилия были затрачены на написание программ, понимающих естественный язык. Хотя такие программы и достигли успеха в ограниченных контекстах, системы, использующие натуральные языки с гибкостью и общностью, характерной для человеческой речи, лежат за пределами сегодняшних методологий.

Понимание естественного языка включает куда больше, чем разбор предложений на индивидуальные части речи и поиск значений слов в словаре. Оно базируется на обширном фоновом знании о предмете беседы и идиомах, используемых в этой области, так же, как и на способности применять общее контекстуальное знание для понимания недомолвок и неясностей, присущих естественной человеческой речи.

Представьте себе, к примеру, трудности в разговоре о футболе с человеком, который ничего не знает об игре, правилах, ее истории и игроках. Способен ли такой человек понять смысл фразы: "в центре Иванов перехватил верхнюю передачу — мяч полетел к штрафной соперника, там за него на "втором этаже" поборолись Петров и Сидоров, после чего был сделан пас на Васина в штрафную, который из-под защитника подъемом пробил точно в дальний угол"? Хотя каждое отдельное слово в этом предложении можно понять, фраза звучит совершенной тарабарщиной для любого нефаната, будь он хоть семи пядей во лбу.

Задача сбора и организации этого фонового знания, чтобы его можно было применить к осмыслению языка, составляет значительную проблему в автоматизации понимания естественного языка. Для ее решения исследователи разработали множество методов структурирования семантических значений, используемых повсеместно в искусственном интеллекте (см. главы 6, 7 и 13).

Из-за огромных объемов знаний, требуемых для понимания естественного языка, большая часть работы ведется в хорошо понимаемых, специализированных проблемных областях. Одной из первых программ, использовавших такую методику "микромира", была программа Винограда SHRDLU — система понимания естественного языка, которая могла "беседовать" о простом взаимном расположении блоков разных форм и цветов [Winograd, 1973]. Программа SHRDLU могла отвечать на вопросы типа: "какого цвета блок на синем кубике?", а также планировать действия вроде "передвинь красную пирамидку на зеленый брусок". Задачи этого рода, включая управление размещением блоков и их описание, на удивление часто всплывали в исследованиях ИИ и получили название проблем "мира блоков".

Несмотря на успехи программы SHRDLU в разговорах о расположении блоков, она была не способна абстрагироваться от мира блоков. Методики представления, использованные в программе, были слишком просты, чтобы передать семантическую организацию более богатых и сложных предметных областей. Основная часть текущих работ в области понимания естественных языков направлена на поиск формализмов представления, которые должны быть достаточно общими, чтобы применяться в широком круге приложений и уметь адаптироваться к специфичной структуре заданной области. Множество разнообразных методик (большинство из которых являются развитием или модификацией *семантических сетей*) исследуются с этой целью и используются при разработке программ, способных понимать естественный язык в ограниченных, но достаточно интересных предметных областях. Наконец, в текущих исследованиях [Marcus, 1980], [Manning и Schutze, 1999], [Jurafsky и Martin, 2000] стохастические модели, описывающие совместное использование слов в языке, применяются для характеристики как синтаксиса, так и семантики. Полное понимание языка на вычислительной основе все же остается далеко за пределами современных возможностей.

1.2.5. Моделирование работы человеческого интеллекта

В большей части рассмотренного выше материала человеческий интеллект служит отправной точкой в создании искусственного, однако это не означает, что программы должны формироваться по образу и подобию человеческого разума. Дейст-

вительно, многие программы ИИ создаются для решения каких-то насущных задач без учета человеческой ментальной архитектуры. Даже экспертные системы, заимствуя бо́льшую часть своего знания у экспертов-людей, не пытаются моделировать внутренние процессы человеческого ума. Если производительность системы — это единственный критерий ее качества, нет особых оснований имитировать человеческие методы принятия решений. Программы, которые используют несвойственные людям подходы, зачастую более успешны, чем их человеческие соперники. Тем не менее конструирование систем, которые бы детально моделировали какой-либо аспект работы интеллекта человека, стало плодотворной областью исследований как в искусственном интеллекте, так и в психологии.

Моделирование работы человеческого разума помимо обеспечения ИИ его основной методологией оказалось мощным средством для формулирования и испытания теорий человеческого познания. Методологии принятия решений, разработанные теоретиками компьютерных наук, дали психологам новую отправную точку для исследования человеческого разума. Вместо того чтобы гадать о теориях познания на неясном языке ранних исследований или вообще оставить попытки описания внутренних механизмов человеческого интеллекта (как предлагают специалисты по изучению поведения), многие психологи приспособили язык и теорию компьютерной науки для разработки моделей человеческого разума. Такие методы не только дают новую терминологию для характеристики человеческого интеллекта. Компьютерная реализация этих теорий предоставляет психологам возможность эмпирически тестировать, критиковать и уточнять их идеи [Luger, 1994]. Обсуждение отношений между ИИ и попытками понять человеческий разум приводится ниже и резюмируется в главе 16.

1.2.6. Планирование и робототехника

Исследования в области планирования начались с попытки сконструировать роботов, которые бы выполняли свои задачи с некоторой степенью гибкости и способностью реагировать на окружающий мир. Планирование предполагает, что робот должен уметь выполнять некоторые элементарные действия. Он пытается найти последовательность таких действий, с помощью которой можно выполнить более сложную задачу, например, двигаться по комнате, заполненной препятствиями.

Планирование по ряду причин является сложной проблемой, не малую роль в этом играет размер пространства возможных последовательностей шагов. Даже очень простой робот способен породить огромное число различных комбинаций элементарных движений. Представьте себе, к примеру, робота, который может передвигаться вперед, назад, влево и вправо, и вообразите, сколькими различными путями он может двигаться по комнате. Представьте также, что в комнате есть препятствия, и что робот должен выбирать путь вокруг них некоторым оптимальным образом. Для написания программы, которая могла бы разумно определить лучший путь из всех вариантов, и не была бы при этом перегружена огромным их числом, потребуются сложные методы для представления пространственного знания и управления перебором в пространстве альтернатив.

Одним из методов, применяемых человеческими существами при планировании, является *иерархическая декомпозиция задачи* (hierarchical problem decomposition). Планируя путешествие в Лондон, вы, скорее всего, займетесь отдельно проблемами организации перелета, поездки до аэропорта, самого полета и поиска подходящего вида транс-

порта в Лондоне, хотя все они являются частью большого общего плана. Каждая из этих задач может сама быть разбита на такие подзадачи, как, например, покупка карты города, преодоление лабиринта линий метро и поиск подходящей пивной. Такой подход не только эффективно ограничивает размер пространства поиска, но и позволяет сохранять часто используемые маршруты для дальнейшего применения.

В то время как люди разрабатывают планы безо всяких усилий, создание компьютерной программы, которая бы занималась тем же — сложная проблема. Казалось бы, такая простая вещь, как разбиение задачи на независимые подзадачи, на самом деле требует изощренных эвристик и обширного знания об области планирования. Не менее сложная проблема — определить, какие планы следует сохранить, и как их обобщить для использования в будущем.

Робот, слепо выполняющий последовательности действий, не реагируя на изменения в своем окружении, или неспособный обнаруживать и исправлять ошибки в своем собственном плане, едва ли может считаться разумным. Зачастую от робота требуют сформировать план, основанный на недостаточной информации, и откорректировать свое поведение по мере его выполнения. Робот может не располагать адекватными сенсорами для того, чтобы обнаружить все препятствия на проектируемом пути. Такой робот должен начать двигаться по комнате, основываясь на "воспринимаемых" им данных, и корректировать свой путь по мере того, как выявляются другие препятствия. Организация планов, позволяющая реагировать на изменение условий окружающей среды, — основная проблема планирования [Lewis и Luger, 2000].

Наконец, робототехника была одной из областей исследований ИИ, породившей множество концепций, лежащих в основе агентно-ориентированного принятия решений (см. подраздел 1.1.4). Исследователи, потерпевшие неудачу при решении проблем, связанных с большими пространствами представлений и разработкой алгоритмов поиска для традиционного планирования, переформулировали задачу в терминах взаимодействия полуавтономных агентов [Agre и Chapman, 1987], [Brooks, 1991*a*]. Каждый агент отвечает за свою часть задания, и общее решение возникает в результате их скоординированных действий. Элементы алгоритмов планирования будут представлены в главах 6, 7 и 14.

Исследования в области планирования сегодня вышли за пределы робототехники, теперь они включают также координацию любых сложных систем задач и целей. Современные планировщики применяются в агентских средах [Nilsson, 1994], а также для управления ускорителями частиц [Klein и др., 1999, 2000].

1.2.7. Языки и среды ИИ

Одним из наиболее важных побочных продуктов исследований ИИ стали достижения в сфере языков программирования и средах разработки программного обеспечения. По множеству причин, включая размеры многих прикладных программ ИИ, важность методологии "создания прототипов", тенденцию алгоритмов поиска порождать чересчур большие пространства и трудности в предсказании поведения эвристических программ, программистам искусственного интеллекта пришлось разработать мощную систему методологий программирования.

Средства программирования включают такие методы структурирования знаний, как объектно-ориентированное программирование и каркасы экспертных систем (они обсуждаются в части III). Высокоуровневые языки, такие как LISP и PROLOG (см. часть IV), которые обеспечивают модульную разработку, помогают управиться с размерами и

сложностью программ. Пакеты средств трассировки позволяют программистам реконструировать выполнение сложного алгоритма и разобраться в сложных структурах эвристического перебора. Без подобных инструментов и методик вряд ли удалось бы построить многие известные системы ИИ.

Многие из этих методик сегодня являются стандартными методами разработки программного обеспечения и мало соотносятся с основами теории ИИ. Другие же, такие как объектно-ориентированное программирование, имеют значительный теоретический и практический интерес. Наконец, многие алгоритмы ИИ сейчас реализуются на таких традиционных для вычислительной техники языках, как C++ и Java.

Языки, разработанные для программирования ИИ, тесно связаны с теоретической структурой этой области. В данной книге рассматривается и LISP, и PROLOG, и мы старались удержаться от религиозных прений об их относительных достоинствах, склоняясь, скорее, к той точке зрения, что "хороший работник должен знать все инструменты". Главы, посвященные языкам программирования (14 и 15), рассматривают преимущества применения различных языков для решения конкретных задач.

1.2.8. Машинное обучение

Обучение остается "крепким орешком" искусственного интеллекта. Важность обучения, тем не менее, несомненна, поскольку эта способность является одной из главных составляющих разумного поведения. Экспертная система может выполнять долгие и трудоемкие вычисления для решения проблем. Но, в отличие от человеческих существ, если дать ей такую же или подобную проблему второй раз, она не "вспомнит" решение. Она каждый раз вновь будет выполнять те же вычисления — едва ли это похоже на разумное поведение.

Большинство экспертных систем ограничены негибкостью их стратегий принятия решений и трудностью модификации больших объемов кода. Очевидное решение этих проблем — заставить программы учиться самим на опыте, аналогиях или примерах.

Хотя обучение является трудной областью, существуют некоторые программы, которые опровергают опасения о ее неприступности. Одной из таких программ является AM — Автоматизированный Математик, разработанный для открытия математических законов [Lenat, 1977, 1982]. Отталкиваясь от заложенных в него понятий и аксиом теории множеств, Математику удалось вывести из них такие важные математические концепции, как мощность множества, целочисленная арифметика и многие результаты теории чисел. AM строил теоремы, модифицируя свою базу знаний, и использовал эвристические методы для поиска наилучших из множества возможных альтернативных теорем. Из недавних результатов можно отметить программу Коттона [Cotton и др., 2000], которая изобретает "интересные" целочисленные последовательности.

К ранним трудам, оказавшим существенное влияние на эту область, относятся исследования Уинстона по выводу таких структурных понятий, как построение "арок" из наборов "мира блоков" [Winston, 1975*a*]. Алгоритм ID3 проявил способности в выделении общих принципов из разных примеров [Quinlan, 1986*a*]. Система Meta-DENDRAL выводит правила интерпретации спектрографических данных в органической химии на примерах информации о веществах с известной структурой. Система Teiresias — интеллектуальный "интерфейс" для экспертных систем — преобразует сообщения на высокоуровневом языке в новые правила своей базы знаний [Davis, 1982]. Программа Hacker строит планы для манипуляций в "мире блоков" посредством итеративного процесса вы-

работки плана, его испытания и коррекции выявленных недостатков [Sussman, 1975]. Работа в сфере обучения, основанного на "пояснениях", продемонстрировала эффективность для обучения априорному знанию [Mitchell и др., 1986], [DeJong и Mooney, 1986]. Сегодня известно также много важных биологических и социологических моделей обучения. Они будут рассмотрены в главах, посвященных коннекционистскому и эмерджентному обучению.

Успешность программ машинного обучения наводит на мысль о существовании универсальных принципов, открытие которых позволило бы конструировать программы, способные обучаться в реальных проблемных областях. Некоторые подходы к обучению будут представлены в главах 9–11.

1.2.9. Альтернативные представления: нейронные сети и генетические алгоритмы

В большей части методик, представленных в этой книге, для реализации интеллекта используются явные представления знаний и тщательно спроектированные алгоритмы перебора. Совершенно отличный подход состоит в построении интеллектуальных программ с использованием моделей, имитирующих структуры нейронов в человеческом мозге или эволюцию разных альтернативных конфигураций, как это делается в генетических алгоритмах и искусственной жизни.

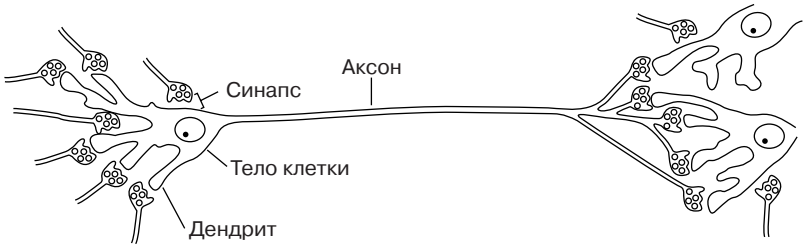

Рис. 1.2. Упрощенная схема нейрона

Схематическое представление нейрона (рис. 1.2) состоит из клетки, которая имеет множество разветвленных отростков, называемых *дендритами*, и одну ветвь — *аксон*. Дендриты принимают сигналы от других нейронов. Когда сумма этих импульсов превышает некоторую границу, нейрон сам возбуждается, и импульс, или "сигнал", проходит по аксону. Разветвления на конце аксона образуют *синапсы* с дендритами других нейронов. Синапс — это точка контакта между нейронами. Синапсы могут быть *возбуждающими* (excitatory) или *тормозящими* (inhibitory), в зависимости от того, увеличивают ли они результирующий сигнал.

Такое описание нейрона необычайно просто, но оно передает основные черты, существенные в нейронных вычислительных моделях. В частности, каждый вычислительный элемент подсчитывает значение некоторой функции своих входов и передает результат к присоединенным к нему элементам сети. Конечные результаты являются следствием параллельной и распределенной обработки в сети, образованной нейронными соединениями и пороговыми значениями.

Нейронные архитектуры привлекательны как средства реализации интеллекта по многим причинам. Традиционные программы ИИ могут быть слишком неустойчивы и чувствительны к шуму. Человеческий интеллект куда более гибок при обработке такой

зашумленной информации, как лицо в затемненной комнате или разговор на шумной вечеринке. Нейронные архитектуры, похоже, более пригодны для сопоставления зашумленных и недостаточных данных, поскольку они хранят знания в виде большого числа мелких элементов, распределенных по сети.

С помощью генетических алгоритмов и методик искусственной жизни мы вырабатываем новые решения проблем из компонентов предыдущих решений. Генетические операторы, такие как скрещивание или мутация, подобно своим эквивалентам в реальном мире, вырабатывают с каждым поколением все лучшие решения. В искусственной жизни новые поколения создаются на основе функции "качества" соседних элементов в прежних поколениях.

И нейронные архитектуры, и генетические алгоритмы дают естественные модели параллельной обработки данных, поскольку каждый нейрон или сегмент решения представляет собой независимый элемент. Гиллис [Hillis, 1985] отметил, что люди быстрее справляются с задачами, когда получают больше информации, в то время как компьютеры, наоборот, замедляют работу. Это замедление происходит за счет увеличения времени последовательного поиска в базе знаний. Архитектура с массовым параллелизмом, например человеческий мозг, не страдает таким недостатком. Наконец, есть нечто очень привлекательное в подходе к проблемам интеллекта с позиций нервной системы или генетики. В конце концов, мозг есть результат эволюции, он проявляет разумное поведение и делает это посредством нейронной архитектуры. Нейронные сети, генетические алгоритмы и искусственная жизнь рассматриваются в главах 10 и 11.

1.2.10. Искусственный интеллект и философия

В разделе 1.1 мы представили философские, математические и социологические истоки искусственного интеллекта. Важно осознавать, что современный ИИ не только наследует эту богатую интеллектуальную традицию, но и делает свой вклад в нее.

Например, поставленный Тьюрингом вопрос о разумности программ отражает наше понимание самой концепции разумности. Что такое разумность, как ее описать? Какова природа знания? Можно ли его представить в устройствах? Что такое навыки? Может ли знание в прикладной области соотносится с навыком принятия решений в этой среде? Как знание о том, *что* есть истина (аристотелевская "теория"), соотносится со знанием, *как* это сделать ("практика")?

Ответы на эти вопросы составляют важную часть работы исследователей и разработчиков ИИ. В научном смысле программы ИИ можно рассматривать как эксперименты. Проект имеет конкретную реализацию в виде программы, и программа выполняется как эксперимент. Разработчики программы изучают результаты, а затем перестраивают программы и вновь ставят эксперимент. Таким образом возможно определить, являются ли наши представления и алгоритмы достаточно хорошими моделями разумного поведения. Ньюэлл и Саймон [Newell и Simon, 1976] предложили этот подход к научному познанию в своей тьюринговской лекции 1976 г.

Ньюэлл и Саймон также предложили более сильную модель интеллекта в своей гипотезе о физической символьной системе: физическая система проявляет разумное поведение тогда и только тогда, когда она является физической символьной системой. В главе 16 подробно рассматривается практический смысл этой теории, а также критические замечания в ее адрес.

Многие применения ИИ подняли глубокие философские вопросы. В каком смысле можно заявить, что компьютер "понимает" фразы естественного языка? Продуцирование и понимание языка требует толкования символов. Недостаточно правильно сформировать строку сим-

волов. Механизм понимания должен уметь приписывать им смысл или интерпретировать символы в зависимости от контекста. Что такое смысл? Что такое интерпретация?

Подобные философские вопросы встают во многих областях применения ИИ, будь то построение экспертных систем или разработка алгоритмов машинного обучения. Эти вопросы будут рассмотрены в данной книге по мере их появления. Философские аспекты ИИ обсуждаются также в главе 16.

1.3. Искусственный интеллект — заключительные замечания

Мы попытались дать определение искусственному интеллекту путем рассмотрения основных областей его исследования и применения. Этот обзор обнаружил молодую и многообещающую область науки, основная цель которой — найти эффективный способ понимания и применения интеллектуального решения проблем, планирования и навыков общения к широкому кругу практических задач. Несмотря на разнообразие проблем, затрагиваемых исследованиями ИИ, во всех отраслях этой сферы наблюдаются некоторые общие черты.

1. Использование компьютеров для доказательства теорем, распознавания образов, обучения и других форм рассуждений.
2. Внимание к проблемам, не поддающимся алгоритмическим решениям. Отсюда — эвристический поиск как основа методики решения задач в ИИ.
3. Принятие решений на основе неточной, недостаточной или плохо определенной информации и применение формализмов представлений, помогающих программисту справляться с этими недостатками.
4. Выделение значительных качественных характеристик ситуации.
5. Попытка решить вопросы семантического смысла, равно как и синтаксической формы.
6. Ответы, которые нельзя отнести к точным или оптимальным, но которые в каком-то смысле “достаточно хороши”. Это результат применения эвристических методов в ситуациях, когда получение оптимальных или точных ответов слишком трудоемко или невозможно вовсе.
7. Использование большого количества специфичных знаний в принятии решений. Это основа экспертных систем.
8. Использование знаний метауровня для более совершенного управления стратегиями принятия решений. Хотя это очень сложная проблема, затронутая лишь несколькими современными системами, она постепенно становится важной областью исследований.

Надеемся, что это введение даст некоторое понятие об общей структуре и значимости искусственного интеллекта. Предполагаем, что краткое обсуждение таких технических вопросов, как перебор и представления, не были излишне поверхностными и неясными, детальнее они будут рассмотрены впоследствии. Здесь они приведены для демонстрации их значимости в общей организации этой области.

Как мы отмечали при обсуждении агентского решения проблем, объекты приобретают смысл при взаимоотношении с другими объектами. Это не менее справедливо в отношении фактов, теорий и методов, образующих любую научную область. Мы по-

пытались дать понятие таких взаимосвязей, так что, когда будут представлены отдельные технические аспекты искусственного интеллекта, они займут свое место в постепенном понимании общей сущности и направлений этой сферы. Мы руководствуемся наблюдением, принадлежащим психологу и системному теоретику Грегори Бэйтсону [Bateson, 1979]:

> "Разрушьте структуру, объединяющую предметы изучения, и вы неизбежно разрушите все его качество".

1.4. Резюме и дополнительная литература

Перспективная область ИИ отражает некоторые из древнейших вопросов, занимавших западную мысль, в свете современного вычислительного моделирования. Понятия рациональности, представления и мышления сейчас находятся, вероятно, под более пристальным рассмотрением, чем во все былые времена, поскольку компьютерная наука требует их алгоритмического понимания! В то же время политические, экономические и этические рамки мира людей заставляют нас помнить об ответственности за последствия наших творений. Надеемся, последующие главы помогут вам лучше понять современные методики ИИ и извечность проблем этой области.

Отличными источниками по вопросам, поднятым в этой главе, являются труды [Haugeland, 1997, 1985], [Dennett, 1978, 1984, 1991, 1995]. Использованы такие первоисточники, как "Физика", "Метафизика" и "Логика" Аристотеля; работы Фреге; труды Бэббиджа, Буля, Рассела и Уайтхеда. Весьма интересны работы Тьюринга, особенно его представления о природе интеллекта и возможности разработки интеллектуальных программ [Turing, 1950]. Биография Тьюринга [Hodges, 1983] представляет собой отличное чтение. Критика теста Тьюринга может быть найдена в [Ford и Hayes, 1995].

Работы [Weizenbaum, 1976] и [Winograd и Flores, 1986] дают трезвую оценку ограничениям и этическим аспектам ИИ. Саймон [Simon, 1981] положительно высказывается об осуществимости искусственного интеллекта и его роли в обществе.

Упомянутые в разделе 1.2 применения ИИ призваны показать широту интересов исследователей ИИ и сформулировать многие изучаемые ныне вопросы. Учебник [Barr и Feigenbaum, 1989] содержит введение в эти области. Кроме того, рекомендуем книги [Nilsson, 1998] и [Pearl, 1984]. В них подробно рассмотрены вопросы ведения игр, которые затронуты в главах 2–5. В главах 2, 3 и 12 обсуждается проблема автоматических рассуждений. Некоторая избранная литература по этой теме включает труды [Wos и др., 1984], [Bledsoe, 1977], [Boyer и Moore, 1979], [Veroff, 1977].

Прочитав главы 6 и 7, читатель может получить хорошее представление об экспертных системах из книг [Harmon и King, 1985], [Hayes-Roth и др., 1984], [Waterman, 1986] и [Durkin, 1994].

Понимание естественных языков — область непрекращающегося изучения, некоторые важные взгляды на нее выражены в [Allen, 1995], [Winograd, 1983], [Schank и Colby, 1973], [Wilks, 1972], [Lakoff и Johnson, 1999], [Jurafsky и Martin, 2000]; введение в эту область представлено в главах 6 и 13 этой книги.

Использование компьютеров для моделирования работы человеческого разума, кратко описанное в главе 16, более подробно обсуждается в [Newell и Simon, 1972], [Pylyshyn, 1984], [Anderson, 1978] и [Luger, 1994]. Планирование и робототехника (см. главы 7 и 14) представлены в т. 3 книги [Barr и Feigenbaum, 1981], [Brooks, 1991*a*] и [Lewis и Luger, 2000].

ИИ-ориентированные языки и среды разработки рассматриваются в главах 14 и 15 этой книги, а также в [Forbus и deKleer, 1993]. Машинное обучение обсуждается в главах 9–11; многотомном издании [Michalski и др., 1983, 1986], [Kondratoff и Michalski, 1990]; журналы *Journal of Artificial Intelligence* и *Journal of Machine Learning* также являются важными источниками информации по этой теме.

Главы 10 и 11 представляют взгляд на интеллект с упором на модульность структуры и адаптацию в социальном и природном контексте. Работа [Minsky, 1985] — одна из наиболее ранних и побуждающих к размышлению работ, отстаивающих эту точку зрения (см. также [Ford и др., 1995] и [Langton, 1995]).

Книга [Shapiro, 1992] дает ясное и всестороннее описание области искусственного интеллекта. Эта энциклопедия также проводит детальный анализ многих разных подходов и противоречий, образующих текущее состояние этой дисциплины.

1.5. Упражнения

1. Предложите и аргументируйте собственное определение искусственного интеллекта.
2. Приведите несколько дополнительных примеров аристотелевского различия между “материей” и “формой”. Можете ли вы показать, как ваши примеры вписываются в теорию абстракции?
3. Западная философская традиция во многом выросла из проблемы взаимоотношения разума и тела. По вашему мнению:

   а) разум и тело — сущности разной природы, каким-то образом взаимодействующие;

   б) разум — всего лишь результат “физических процессов”;

   в) тело — лишь иллюзия мысли?

   Обсудите ваши мысли по поводу проблемы разума и тела и ее важность для теории искусственного интеллекта.
4. Приведите критические замечания по поводу тьюринговского критерия “разумности” компьютерной программы.
5. Сформулируйте ваш собственный критерий “разумности” компьютерной программы.
6. Хотя вычислительная техника — относительно молодая дисциплина, философы и математики размышляли о вопросах, существенных для автоматического решения задач, на протяжении тысяч лет. Каково ваше мнение о существенности этих философских вопросов по отношению к проектированию устройств интеллектуального решения проблем? Аргументируйте свой ответ.
7. Рассмотрев различия между архитектурами современных компьютеров и человеческого мозга, поясните, какое значение имеет исследование физиологической структуры и функции биологических систем для разработки программ ИИ? Аргументируйте свой ответ.
8. Выберите проблемную область, в которой, как вы считаете, затраты на разработку экспертной системы были бы оправданы. Разъясните в общих чертах суть проблемы. На основании своей интуиции скажите: какие аспекты принятия решений будет наиболее сложно автоматизировать?

9. Найдите еще два преимущества экспертных систем, кроме уже перечисленных в тексте. Обсудите их с точки зрения интеллектуальных, социальных или экономических результатов.
10. Поясните, почему вы считаете такой сложной проблему машинного обучения.
11. Считаете ли вы возможным для компьютера понимать и использовать естественный (человеческий) язык?
12. Выявите и поясните два потенциально негативных последствия развития искусственного интеллекта для общества.