第1章 绪 论

我认为，理解智能包括理解：知识如何获取、表达和存储；智能行为如何产生和学习；动机、情感和优先权如何发展和运用；传感器信号如何转换成各种符号；怎样利用各种符号执行逻辑运算、对过去进行推理及对未来进行规划；智能机制如何产生幻觉、信念、希望、畏惧、梦幻甚至善良和爱情等现象。我相信，对上述内容有一个根本的理解将会成为与拥有原子物理、相对论和分子遗传学等级相当的科学成就。

— James Albus“答复 Henry Hexmoor”，摘自URL：
http://tommy.jsc.nasa.gov/er/er6/mrl/papers/symposium/albus.txt
1995年2月13日

1.1 什么是人工智能

广义地讲，人工智能是关于人造物的智能行为，而智能行为包括知觉、推理、学习、交流和在复杂环境中的行为。人工智能的一个长期目标是发明出可以像人类一样或能更好地完成以上行为的机器；另一个目标是理解这种智能行为是否存在于机器、人类或其他动物中。因此，人工智能包含了科学和工程的双重目标。本书主要从工程角度讨论 AI，集中说明构成智能机器设计基础的重要概念和思想。

长期以来，围绕着人工智能有很多争议。“机器是否能思考？”这一问题吸引了许多哲学家、科学家和工程师。在一篇著名的文章中，计算机科学的创始人之一，艾仑·图灵（Alan Turing），重述了这一问题，使其更经得起一种实验的测试，这种测试后来被称为图灵测试 [Turing 1950]。下面将描述这一测试，但图灵同时指出对“机器是否能思考”这一问题的答案取决于人们如何定义“机器”和“思考”。他也许还可指出，这一问题还依赖于人们如何定义“能”。

让我们先来考虑“能”这个词。我们认为机器现在或将来能思考吗？我们认为原则上机器应该可以思考吗（即使我们不可能制造出这样的机器）？或者，我们真的要求实际的演示吗？由于人造物尚未具有广泛的思考能力，这些问题就变得非常重要。

一些人认为，能够思考的机器必定十分复杂且拥有复杂的经验（如与其所处的环境和其他能够思考的机器交流）。以致于我们永远也无法设计并制造出它们。产生全球气象的过程是一个很好的例子。尽管我们知道有关天气的一切重要现象，这些知识也无法让我们完整、详尽地复制天气现象。因为再没有比地球表层、大气层和海洋这些存在于宇宙之中、汲取太阳的光和热并受潮汐影响的更简单的系统能够完整详尽地复制天气现象了。同样，完全与人类相当的智能会十分复杂，或者至少会十分依赖于人类严密的生理机能，从而使其不能脱离处于特定环境的人的主体（*embodiment*）而单独存在（关于“主体”这一概念的重要性的讨论，可参见 [Lakoff 1987， Winograd & Flores 1986， Harnad 1990， Mataric 1997]）。至于我们是否能造出与人类水平相当的能思考的机器仍无定论。但人工智能朝着这一目标的发展是坚定不移的，虽然这一进展比早期开创者们的预计要慢。我对我们最终的胜利持乐观态度。

接着，我们考虑“机器”这一词。许多人认为，机器是一种相当愚钝的东西，它总让人联

想起齿轮转动、蒸汽嘶嘶、钢铁锵锵的景象。这样的机器能思考吗？但是，如今计算机已大大延伸了“机器”这一概念。同时，我们对生物机制的理解也有了前所未有的进展。譬如：一种名为E6抗菌素的简单过滤性病毒（如图 1-1所示），它头部含过滤性病毒 DNA。它用尾部须根与一个细菌的细胞壁相连，先刺入细胞壁，再将其 DNA注入此细菌中，然后这些 DNA使此细菌产生成千上万这一过滤性病毒 DNA的复制品。这些复制品自动集合而形成新的过滤性病毒后，离开这一细菌，再重复以上过程。这一完整的集合看起来、运作起来均像一台机器，我们还不如称之为由蛋白质构成的机器。

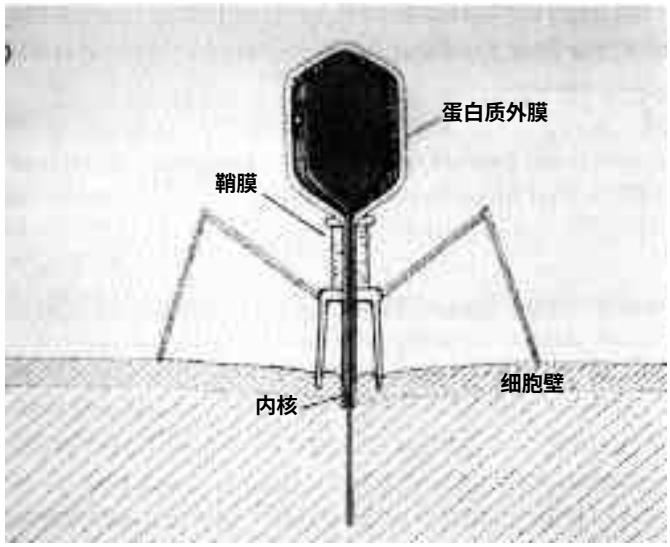

图1-1 E6抗菌素的示意图

其他生物过程和有机物又如何呢？最近，细菌 Haemophilus influenzae Rd的基因排列已经完全被破解[Fleischmann, et al. 1995]。这一基因有1 830 137个基对（由字母A、G、C和T组成），大概占3.6×10^6比特，约半兆字节。尽管科学家们还不清楚其中 1743个基因的功能，但他们已经开始像解释机器(当然是非常复杂的机器)的发展和功能那样解释这些机制的发展和功能。运用逻辑电路时序图这一计算机科学家熟知的技术，对理解可感染细菌的过滤性病毒的复杂生化基因的规则是十分有益的[McAdams & Shapiro 1995]。对其他有机物包括人类的基因组的排序工作正在进行中。一旦知道这些结果，我们会把细菌、寄生虫、果蝇、老鼠、海豚与人类一起视为机器吗？如果人类是机器，那么机器便能思考！因为我们拥有活生生的证据。只是我们还不知道人类这种机器是如何运作的。

即使我们就什么是“机器”达成一致观点，这种观点仍存在其他许多争议。尽管由蛋白质构成的机器能思考，由硅片构成的机器则未必能。一位知名哲学家 John Searle认为，我们由什么构成直接影响着我们的智能[Searle 1980, Searle 1992]。他认为思考仅发生在那些十分特殊的机器上——有生命且由蛋白质构成的机器。

与Searle的观念（和前面所提到的“主体”概念）截然相反，Newell 和 Simon提出了物理符号系统假说[Newell & Simon 1976]。这一假说指出，物理符号系统具备必要且足够的方法来进行普通智能行为。Newell 和 Simon指出物理符号系统是类似数字计算机的机器，具备灵活处理符号数据的能力——加数、重排符号序列（如按字母顺序排列一组姓名）及符号替换等等。

这一假说的重要之一是它指出这种物理符号系统由什么构成并不重要！这一假说是完全中性的。一个智能实体只要能处理符号，它可以由蛋白质、机械传动、半导体或其他什么构成[1]。

还有一些人也认为，机器是由蛋白质还是由硅片构成无关紧要，但他们认为智能行为是他们所谓的“亚符号处理”，即“信号处理”，而不是“符号处理”的结果。如识别熟悉面孔对人类来说易如反掌，而我们却不知道机器该如何运作。他们认为这一过程最好的解释便是人类把图象或图象各部分作为多维信号而不是符号来处理。

能列出很多其他关于什么样的机器才具有人类的思维能力的看法，我们经常听到的有：

- 人脑对信息进行并行处理，而传统的计算机则是串行处理。我们需建造各种新型的并行计算机来加快人工智能的发展。
- 传统的计算机以非真即假（双态）逻辑为基础，而真正的智能系统应运用某种模糊逻辑。
- 动物神经元远比开关——计算机的基本模块——要复杂，我们需要在智能机器中运用更现实的人造神经元。

尽管许多人工智能研究者接受物理符号系统这一假说，但在人工智能领域关于究竟需要哪种机器达成共识还为时过早。

最后，我们来看看“思考”这一最难的词。图灵没有企图对这个词下定义，只是提出了“图灵测试”。通过这一测试即可判断某一特定机器是不是智能机器。这一测试最初被描绘成一种游戏。从图灵的文章中摘录如下[Turing 1950]：

> 游戏由一男（A）、一女（B）和一名询问者（C）（性别不限）进行。询问者单独在一间房间里与其他两人分别通过电传打字机联系。在游戏中，询问者的目的是分辨两人的性别。开始，他只知道两人的称呼X、Y，最终，他需要在“X是A，Y是B”或者“X是B，Y是A”中选择答案。询问者允许问A和B以下问题：
>
> C：X能告诉我你的头发的长度吗？
>
> 如果X是A，那么他必须回答。游戏中，A必须尽力使C判断错误。
>
> ···
>
> 而B的任务则是帮助询问者。
>
> ···
>
> 现在我们提出这样一个问题：一个机器代替游戏中的A会如何？询问者会依然像当游戏由一男一女进行时一样经常判断错误吗？这些问题代替了最初的问题：机器能思考吗？

图灵测试常被简化为让一个机器试图使询问者相信它是一个人。许多更简单的测试层出不穷，然而由于就连一些陈旧的机器也可以愚弄询问者一段时间，这些简单的测试已经不再被视为测试机器智能的良方了。譬如：Joseph Weizenbaum的ELIZA程序运用一些相当简单、但对一个宽容的使用者却是虽显空洞却十分现实的对话技巧。Mauldin的JULIA程序是更新和更复杂的对话程序[Mauldin 1994][2]。

除了运用图灵测试，我们有必要在标榜一台机器是智能机器之前，了解这样的机器应具备

---

[1] 当然，如果我们考虑到速度、永久性、可靠性、并行处理的适合性和温度敏感度等实际因素，模块材料必定有好有坏。

[2] 1991年，Hugh Loebner开始举行一个有奖竞赛，他向第一个能通过无限制图灵测试的计算机程序的开发者提供10万美元的奖金。另外，每年这一竞赛都为能通过有限图灵测试的最佳程序的开发者提供数额略少的奖金。

怎样的能力。许多计算机程序已经完成了大量不可思议的事——设计高效省油的最佳航空路线、模拟全球气象状况、统筹安排工厂的机器使用等等。这些是智能程序吗？它们能体现人工智能的主旨吗？本书一开始我便描绘那些难以被人们称为智能机器的机器，随着其复杂性的增强，它们会变得越来越智能吗？毫无疑问，别人会有不同的观点，但至少我这样认为。

1.2 人工智能的研究方法

尽管人工智能已经创造了一些实用系统，但人们不得不承认这些远未达到人类的智能水平。正因为如此，就选择人工智能研究的最佳方法——既为人工智能的最终研究目标打好基础，又能创造出短期效益——存在大量的讨论和争辩。这样，在过去的四十年里涌现出大量方法，每一种方法均有其拥护者，有些甚至有趣得令人爱不释手。也许所有这些方法应该综合起来运用。总之，所有这些拥护者都认为自己的研究方法具有突破性进展，值得特别关注。其中的一些方法可分为两大类。

第一类包括符号处理的方法。它们基于Newell和Simon的物理符号系统的假说。尽管不是所有人都赞同这一假说，但几乎大多数被称为“经典的人工智能”（即哲学家John Haugeland所谓的“出色的老式人工智能”或GOFAI）均在其指导之下。这类方法中，突出的方法是将逻辑操作应用于说明性知识库。最早由John McCarthy的“采纳意见者”备忘录提出[McCarthy 1958]，这种风格的人工智能运用说明语句来表达问题域的“知识”，这些语句基于或实质上等同于一阶逻辑中的语句。采用逻辑推理可推导这种知识的结果。这种方法有许多变形，包括那些强调对逻辑语言中定义域的形式公理化的角色的变形。当遇到“真正的问题”，这一方法需要掌握问题域的足够知识，通常就称作基于知识的方法。许多系统的构建都运用了这些方法，在本书后面将会提到一些。

在大多数符号处理方法中，对需求行为的分析和为完成这一行为所做的机器合成要经过几个阶段。最高阶段是知识阶段，机器所需知识在这里说明。接下来是符号阶段，知识在这里以符号组织表示（例如列表可用列表处理语言LISP来描述），同时在这里说明这些组织的操作。接着，在更低级的阶段里实施符号处理。多数符号处理采用自上而下的设计方法，从知识阶段向下到符号和实施阶段。

第二类包括所谓的“子符号”方法。它们通常采用自下而上的方式，从最低阶段向上进行。在最低层阶段，符号的概念就不如信号这一概念确切了。在子符号方法中突出的方法是“*Animat approach*”。偏爱这种方式[Wilson 1991，Brooks 1990]的人们指出，人的智能经过了在地球上十亿年或更长时间的进化过程。他们认为，为了制造出真正的智能机器，我们必须沿着这些进化的步骤走。因此，我们必须集中研究复制信号处理的能力和简单动物如昆虫的支配系统，沿着进化的阶梯向上进行。这一方案不仅能在短期内创造实用的人造物，又能为更高级智能的建立打好坚实的基础。

第二类方法也强调符号基础。[Brooks 1990]将物理符号系统和他的物理基础假说相对照。在物理基础假说中，一个agent不采用集中式的模式而运用其不同的行为模块与环境相互作用来进行复杂的行为（然而，他也承认，要达到人类智能水平的人工智能也许需要将两种途径相结合）。

机器与环境的相互作用产生了所谓的“自然行为（*emergent behavior*）”。一名研究人员这样说[Maes 1990b，p.1]：

一个agent的功能可视作该系统与动态环境密切相互作用的自然属性。agent本身对其行为的说明并不能解释它运行时所表现的功能；相反，其功能很大程度上取决于环境的特性。不仅要动态地考虑环境，而且环境的具体特征也要运用于整个系统之中。

由子符号派制造的著名样品机器包括所谓的“神经网络（*Neural network*）”。受到生物学方法的启发，这些系统主要因其学习的能力而十分有趣。根据模拟生物进化方面的进程，一些有趣的机器应运而生，包括：Sexual crossover、Mutation和Fitness-proportional reproduction。其他自下而上、含animat风格的方法是基于控制理论和动态系统的分析（参见[Beer 1995，Port & van Gelder 1995]）。

介于自上而下和自下而上之间的方法是一种动机“环境自动机（*situated automata*）”[Kaelbling & Rosenschein 1990，Rosenschein & Kaelbling 1995]的方法。Kaelbling 和 Rosenschein 建议编写一种程序设计语言来说明agent在高水平上所要求的行为，并编写一编译程序，以从这种语言编写的程序中产生引发行为的线路。

1.3 人工智能简史

当20世纪40~50年代数字计算机研制成功时，几位研究者就编写了能够完成原始推理工作的程序。其中突出的是第一个可以下国际象棋[Shannon 1950， Newell，Shaw & Simon 1958]、担当实验员[Samuel 1959, Samuel 1967]和证明平面几何定理[Gelernter 1959]的计算机程序。1956年，John McCarthy和Claude Shannon合作编著了一本名为《Automata Studies》（自动机研究）的书[Shannon & McCarthy 1956]。由于对书中主要针对automata的数学理论感到遗憾，所以McCarthy决定把1956年的Dartmouth会议用人工智能来命名。在该次会议上发表了许多重要论文，包括由Allen Newell、Cliff Shaw和Herbert Simon 编写的名为《Logic Theorist》（逻辑理论家）[Newell, Shaw & Simon 1957]的程序，它可以证明命题逻辑中的定理。尽管人们试着用许多其他名称来为该领域命名，包括复杂信息处理、机器智能、启发式编程和认知技术，但人工智能这一名称最终保留下来。毫无疑问，这主要归因于一系列的教科书、大学课程、会议和期刊均用这一命名。

很久以前，亚里士多德（公元前384~322年）在着手解释和编纂他称之为三段论的演绎推理时就迈出了向人工智能发展的早期步伐。一些使智能自动化的努力对于今天来说显得太不实际。一位加泰罗尼亚的诗人兼神秘主义者，Ramon Llull（大约1235~1316年），构建了一套称为Ars Mgna的转轮，据说是一部可以回答任何问题的机器。同时许多科学家和数学家开始探讨推理自动化。Martin Gardner[Gardner 1982，p. 3]把“有一天所有的知识，包括精神和无形的真理，能够通过通用的代数演算放入一个单一的演绎系统”的梦想归功于莱布尼兹（1646~ 1716年）。莱布尼兹称这个系统为微积分原理机，或推理机。当然，这个梦想运用当时的技术设备是无法实现的。直到布尔[Boole 1854]建立并发展了命题逻辑，这方面才有了实质性的进展。布尔的意图是要“把有关人类意识的本质和构成的某些可能的暗示收集起来”。到了19世纪末期，Grottlieb Frege提出了用于机械推理的符号表示系统，从而发明了我们现在熟知的谓词演算[Frege 1879]，他称之为Begriffsschrift，可以译为“概念书写（*concetp - writing*）”。

1958年，John McCarthy 建议在他称之为“意见采纳者”的系统中采用谓词演算这种语言来表示和运用知识。这一系统被告知它所需要知道的而不是事先程序设计好的知识。 Covdell

Green在他所谓的QA3系统中[Green 1969a]适度地、颇有影响地实现了这些思想。尽管在研究者中还存在大量争议，谓词演算和一些它的变形构成了人工智能知识表示的基础。

20世纪的逻辑学家，包括Kurt Cödel、Stephen Kleene、Emil Post、Alonzo Church和Alan Turing，对哪些能和哪些不能由逻辑和计算机系统完成的任务做了形式化分类。最近，计算机科学家，包括Stephen Cook和Richard Karp证明有些计算在原则上可能需要根本不切实际的时间和存储空间。

许多从逻辑学和计算机科学中所得到的结果是：“真理不可能被演绎”、“计算不可能被执行”。也许这些负面的发现令许多哲学家和其他人振奋，他们将之理解为再一次否定了人类的智能可以机械化[Lucas 1961, Penrose 1989, Penrose 1994]，他们猜想人类不存在机械所固有的计算局限。然而多数逻辑学家和计算机科学家却认为这些负面结果并不暗示机器具有人类所不具有的任何局限。

在现代，第一篇讨论把人类智能机械化的可能性的文章是由 Alan Turing所著的（前面已经引用）[Turing 1950]。同一时期，Warren McCulloch和Walter Pitts总结出简单计算元素和生物神经元之间关系的理论[McCulloch & Pitts 1943]。他们证明了运用逻辑网络系统计算可计算功能的可能性（参见 [Minsky 1967]有关McCulloch Pitts神经元计算方面有价值的论述）。另外，由Frank Rosenblatt[Rosenblatt 1962]所著的书中探讨了称作perceptrons的网络由类似于神经元的部件组成运用于学习和模式识别的可行性。一些其他学派的工作，如控制论 [Wiener 1948]、认知心理学、计算语言学[Chomsky 1965]和自适应控制理论[Widrow & Hoff 1960]，均对人工智能的发展作出了贡献。

许多人工智能的早期工作（从20世纪60年代至70年代初）探讨了问题表示、搜索技术和通用启发等一系列问题——并把它们运用于计算机程序中来解迷、博奕和检索信息，其中有影响的程序是由Allen Newell、Cliff Shaw和Herbert Simon[Newell，Shaw & Simon 1959，Newell & Simon 1963]共同编写的通用问题求解程序（General Problem Solver（GPS））。由这些早期系统解决的实例问题包括符号集成 [Slagle 1963]、代数词汇问题[Bobrow 1968]、类比难题[Evans 1968]及机器人的控制[Nilsson 1984b]。在这些系统中，许多都是《Computers and Thought》这卷书中的主题[Feigenbaum & Feldman 1963]。

为了应用于更重要的现实问题而对这些程序和技术进行升级的尝试表明这些系统只能解决“玩具问题”。更有效的系统要求对应用领域具有更多内在的知识。20世纪70年代末80年代初发展了一些更高级的程序，包括在完成一定任务时模拟专业人员的知识，如分析、设计和诊断等。一些表达具体问题的知识得到了探讨和发展。第一个能演示具体领域知识的重要程序DENDRAL是一个根据所提供的化学分子式和质谱分析图来预测有机物分子结构的系统[Feigenbaum，Buchanan & Lederberg 1971，Lindsay, et al. 1980]。接着，其他“专家系统”，包括医疗诊断[Shortliffe 1976，Miller，Pople & Myers 1982]、计算机系统的配置[McDermott 1982]和矿藏评估（evaluated potential ore deposits）[Campbell，et al. 1982，Duda，Gaschnig & Hart 1979]。[McCorduck 1979]撰写了这一阶段的人工智能简史。

通过升级游戏问题，博奕这一领域有了实质性的进展。 1997年5月11日，一个名为“深蓝”的IBM程序在六局比赛中以 3.5比2.5的总比分战胜了世界象棋冠军 Garry Kasparov(盖利·卡斯帕洛夫)。这次成功是运用复杂的搜索算法、高速计算机和国际象棋专用硬件才得以实现的。

人类的智能包括洞察和分析可视场景、理解并运用语言等许多方面的能力。关于这些能力的专题均得到了高度重视。Larry Roberts开发了早期场景分析程序之一[Roberts 1963]。这一工作之后对机器视觉作了大量研究（[Nalwa 1993]是一本很好的通用教材），对动物视觉系统[Letvinn，et al. 1959，Hubel 1988，Marr 1982]也作了研究。一个早期自然语言理解系统也由Terry Winograd开发成功[Winograd 1972]。20世纪70年代，在一个多方项目中，连续语言理解系统原型被开发出来；由William Woods开发的LUNAR系统能回答用口语提出的关于由美国航空航天局（NASA）从月球收集的岩石样品的问题。尽管现在存在一些自然语言理解系统，但它们的能力仅局限于特定的话题和词汇。打破这些局限有待于开发出更大量的常识表示。CYC项目[Guha & Lenat 1990，Lenat & Guha 1990，Lenat 1995]的一个目标就是尽可能多地收集、表达这些所需的知识。

20世纪50年代末在Frank Rosenblatt所领导的开创性工作之后，对神经网络的研究虽然一度萎靡，但是到20世纪80年代又恢复了活力。具有强度可调互连系统的非线性元素网络如今已被视为一类重要的非线性建模工具。现在已存在神经网络方面的一些重要应用。动态方法与神经网络相结合，促使人工智能的研究集中到把符号处理过程与处于物理环境中的机器人的传感器与受动器联系起来的问题上来。

立足当前，展望未来，我认为人们将重视集成的、自治的系统——机器人和Softbots。Softbots[Etzioni & Weld 1994]是在互联网中查找他们认为用户会感兴趣的信息的软件agent。今后，不断提高和完善机器人和软件agent的能力将促进并引导人工智能研究。

1.4 本书规划

许多人工智能研究者已提出了一些有关智能机械化的观点和技术，我会在介绍一系列逐步弹大逐步复杂的agent时陈述这些内容。我们本可以考虑各类agent及其环境，如：在太空失重的情况下、在海底深水域中、在办公楼或工厂中及在互联网的“符号数据世界”中的机器人。然而，在这样的真实世界中，真正实用的agent将会十分复杂，这样会难以清晰地展示赋予agent智能的人工智能概念。因此，我将在“网格空间世界”这一假想空间中采用一系列“玩具”agent。虽然简单世界很容易描述，但各方面的发展使之变得太复杂而迫使其中的agent需要有智能才行。

网格空间世界是一个三维空间，它以二维的地面为界限，而地面是由一系列单元格组成。单元格集合可以容纳具有各种特性的物体。单元格集合之间可能会存在像墙一样的边界，agent不能离开地面，但可以在单元格之间移动。物体必须在地面上或必须由在地面上的其他物体支撑。有时我会采用仅包含地面的二维子空间。一个典型的网格空间世界如1-2图所示。图中有两个机器人，一个是原始的二维机器人，它用感知邻近单元格是否为空的传感器来判定是否向其移动；另一个略微复杂，它有一个可操纵物体的手臂。

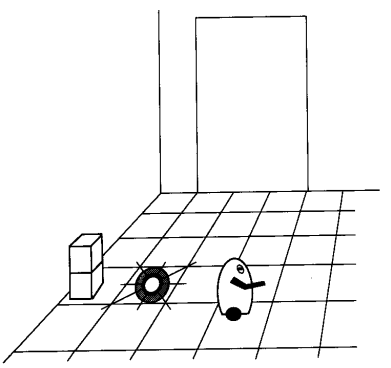

图1-2 网格空间世界

熟悉人工智能历史的读者会发现网格空间世界能够被定制为其他许多用于人工智能研究

的“世界”，包括积木世界、瓷砖世界［Pollack & Ringuette 1990］、wumpus世界［Russell & Norvig 1995，pp. 153以后］和蚂蚁世界［Koza 1992，pp. 54以后］。这些世界仅包含有限的位置、agent、物体和时间点。从这一意义上讲，它们均是离散的。我所描述的多数人工智能技术仅适用于离散世界，而且需要用子符号处理，使它们与连续世界相连。

本书的开头介绍响应agent，它们运用不同方式感知世界并活动于其中。更复杂的agent具有记忆世界特性并存储世界内部模型的能力。任何情况下，响应agent的行为都是所感知和记忆的世界的过去或现在状态的函数。它们能进行相当复杂的知觉和驱动处理。虽然在第6章将详细介绍视觉感知，但因篇幅所限，有关机器人驱动的底层控制的内容无法囊括其中。

多数人工智能系统对它们所处的世界和任务采用某种模型和表示。从广义上讲，模型是与世界紧密相关的任何符号结构和计算的集合，在此基础上计可算产生agent所需的有关世界的信息。这一信息可能是当时agent所处世界或今后可能所处世界的状态。在由浅入深地介绍人工智能agent的过程中，我把人工智能系统分为两种模型。一种是“图标”模型，包括模仿agent环境各方面以及agent的行为对环境产生的影响的数据结构和计算。在一张8×8单元格数组中，表示棋局状况便是图标表示的一例。如果这一表达包括所有棋子位置的信息，则这一棋局图标模型就是完整的。

另一种是“基于特征”的模型。这种模型对环境进行描述。譬如说在棋局中有两个特征，一个特征可能是车或王是否安全，另一个可能是王被将了几次军。描述特征集往往不完整——这是基于特征的表示方式的优点，它们能容忍agent对复杂世界的知识了解的残缺。

接下来介绍的一系列agent能够预计其行为的效果，并采取那些能够达到预期目的的行动。这样的agent可以说具备规划能力。一些研究者认为，这种能力是判断机器是否具备智能的标准，而人工智能也应该由此开始发展。在不能完全感知和模型化的世界中采取行动的agent也需要时刻了解它们的行为是否达到预期效果。

多数网格空间世界具有与真实世界的特性类似的隐含约束条件。如一个物体已有一个具体位置，在同一时刻它就不可能有另一个位置。能够考虑这些约束条件的agent通常更有效力。然后，我将介绍一系列具有推理能力的agent，它们能演绎出隐含于约束条件中的、所处世界的特性。

最后，我将介绍已有其他agent占据的世界中的agent。它们出色的表现有时依赖于对其他agent行为的预期和影响。agent之间的交流十分重要。

在不断加大agent复杂程度的过程中，我总是会讨论agent学习它们所处环境的方法。除了计划能力，学习能力也被视为一个智能系统品质的证明，这一点，我与[Russell & Wefald 1991，p. 18]的观点一致。Russell和Wefald写到：

> 学习是自治性的一个重要方面。一个系统可称之为自治的，即它的行为是由其自身的当前输入和过去的经验而不是设计者的输入和经验来决定。agent往往为一类环境专门设计，这类环境中的每一种情况都已经存储在agent中，并且与设计者所了解的真正环境相一致。这样一个在固有的假设基础上操作的系统只有当这些假设完全真实时才可能成功运行，因而缺乏灵活性。如果给予充分的应变时间，一个真正自治的系统应能在任何环境中成功地运行。原则上，这一系统的内部知识结构应该可以根据其自身对世界的经验而进行构造。然而不能把自治系统与tabula rasa系统等同起来。一个合理折衷的观点是在一开始根据设计者对世界的知识来设计大部分系统的行为，但所有这些假设必须尽可能明确且易

于为agent 所修改。这种意义的自治和我们最初关于智能的概念也完全吻合。

除了用来统一集中论述人工智能技术而使用的网格空间，我还会不断地介绍一些解决现实问题的重要应用。

1.5 补充读物和讨论

关于反对使用图灵测试的争论（不幸的是这似乎也是放弃人工智能达到人类智能水平的宏伟目标的争论），请参阅[Hayes & Ford 1995]，但图灵测试在小说《Galetea 2.2》[Powers 1995]中却十分有趣。

人工智能的自上而下和自下而上的两种研究方法均受到对人类和动物行为研究的启发。进行自下而上研究的人倾向于集中对可以通过组织类神经元的计算元素或逻辑门而实现的行为（通常是简单的行为）进行研究。像动物行为学家一样，他们借此创造了动物行为的各种计算模型。本章提到的许多所谓基于行为的动态人工智能的研究方式是从动物模型中得来的。就动物和机器人之间的比较的讨论，请参阅 [Anderson & Donath 1990，Beer，Chiel & Sterling 1990]。

像心理学家那样，自下而上的研究者和神经科学家也已经发展了试图解释某些人类知觉和行为的神经网络模型。这样的神经网络系统有的可以学会朗读手写体语句 [Sejnowski & Rosenberg 1987]，有的可以识别尺寸、方位和姿态不同的字母数字字符 [Minnix，McVey & Iñigo 1991]，有的具有在书信识别方面上下文敏感的洞察力 [McClelland & Rumelhart 1981，McClelland & Rumelhart 1982]。

当自上而下的研究者从动物和人类的行为中获取灵感时，他们倾向于将研究的焦点集中在那些可用符号来处理最佳模型化的领域，包括认知心理学家研究的问题求解、语言和记忆任务。开发人类问题求解的计算机模型的两位先驱是 Herbert Simon 和Allen Newell（参阅[Newell & Simon 1972，Newell 1991]）。至于后一本书的评论和 Newell所做的回答，请参阅 [Artificial Intelligence, vol 59, 1993]。而关于认知科学和计算机科学之间的关系，请参阅 [Johnson-Laird 1988]。

当然，你也可以认为无论动物和人类如何完成智能行为与设计智能人造物这一工程问题毫无关系，就如同飞机并不像鸟或昆虫那样飞行。设计优良的智能机器，尽管也许能够超越人类，但也许与自然发生的智能行为根本不同。通常引用“蛮力（*Brute-force*）”搜索(因此假定为非人类所为)方式在许多领域的成功应用，包括：玩游戏、列时间表和规划， [Ginsberg 1996]推测机器更高级的思考方式也许会与人类和动物的大脑截然不同，以至拘泥于这种体系结构的人工智能将无法复制人和动物的行为[Dreyfus 1979，Dreyfus 1992，Dreyfus & Dreyfus 1986]。

其他学科对人工智能的影响并不总是定位于自下而上或自上而下的方式。譬如说，[McFarland & Bösser 1993]讨论了许多发生在动物身上的计算实例，然而他们的观点却与经济学和公用事业理论紧密相连。他们认为，动物是经济 agent。本书第3章将着重讨论公用事业理论在动物模型中的作用。一个名为 Michael Wellman的人工智能研究者发明了一种称为“面向市场的程序设计”方法[Wellman 1996]。[Shoham 1996]，也陈述了相关观点。

如何定义智能行为正是区别自下而上和自上而下两种方式的根本所在。不管一种行为是如何“计算”出来，只要它是正确的，就能称之为智能行为吗？实际上存在两种相反的意见。[McFarland & Bösser 1993, p. 6]认为“智能行为是能够得出正确答案的行为，而不管这一答案

是如何得出的”；相反，[Russell & Wefald 1991, p. 1]认为，“……由于任一物理系统在推理能力方面不可避免的局限使其不可能在任何时候都能做出正确的举动”，“智能系统的设计者需要忽略其举动的正确性而去考虑设计正确的系统……”。这一观点引出了基于“有限合理性”的方法。他们主张智能系统必定是在两种行为中决策：一种是在世界中的行为；另一种是旨在完善对其所在世界中最佳行为的估计的计算行为（请参阅[Russell 1997]）。

可计算性的确限制了智能系统所能完成的任务。根据复杂性分析[Garey & Johnson 1979]的原则，人们更加注重对“易处理的”（即可在多项式空间与时间代价下实现的）任务的计算。然而复杂性分析通常涉及最坏情况（而不是平均情况）的结果，许多有趣的人工智能计算都具有最坏情况的指数计算复杂度。我想这样回答那些对许多人工智能算法的不易处理性忧心忡忡的人比较合适，即我们寻找具有平均情况表现良好的算法，并且在许多情况下愿意寻找粗略的、非最佳的解决方法。

对日益丰富的agent的发展的详细说明与动物智能进化的某些步骤息息相关，这对我来说并不新鲜。[Dennett 1995, pp 373以后]提出了一个相似的agent的进化顺序，他将之命名为Darwinian，Skinnerian，Popperian和Gregorian。

尽管我们仍远远无法创造出具有一般人类智能的系统，但想一想这样做的结果却十分重要。当然，廉价的机器人、Softbot、自然语言系统和专家系统具有相当的经济价值。运用这些系统会导致大量失业吗？或者会像早期的工业技术那样创造出比淘汰的更多的就业机会吗？如果说多数这些工作能够由它们这些智能系统完成又会如何呢？（至于我对这些问题的早期想法，请参阅[Nilsson 1984a]）。Joseph Weizenbaum[Weizenbaum 1976]曾担心另外一个问题，即将这些人工智能系统运用到他认为不适当的任务（如辩护、教学和审判）中去的危险。因为自动系统往往造成一种假像，即它们能完成对它们来说实际上根本无法完成的任务，所以会出现对人工智能系统一味草率依赖的危险。然而对这些稍有弥补的是，自动系统比人类少出错。至于人工智能系统对各方面产生的影响的文章，请参阅[Trappl 1986]。

也许，人工智能对我们理解人类自身所产生的影响最深刻。哥白尼和其他天文学家把我们从宇宙的中心带到了不计其数的银河系中的一个小行星上；达尔文和其后的进化论者又使我们从天地万物的中心变成了现在的基于DNA的不计其数的生命形态之一。我们过去都曾难以接受所有这些观念的变化，那么，如果我们真的成功造出像我们一样聪明的机器，我们又将面临什么呢？

本节末尾，我简要列出了有关人工智能的资料出处，它们会随本书后面章节中介绍的子论题的增多而增多。一些重要的期刊包括：《Artificial Intelligence》、网上的《Journal of Artificial Intelligence Research》、《Computational Intelligence》和《Journal of Experimental and Theoretical Artificial Intelligence》。重要的会议包括：年度国家人工智能会议（由AAAI主办），两年一次的国际人工智能联合会议（IJCAI）。一些国家和地区举办的会议还发表会议论文集，如欧洲人工智能会议（ECAI）。AAAI还举办年度春季专题讨论会和秋季研讨会，借此宣布及讨论最新的研究成果。计算机械协会（ACM）还拥有一个特殊的人工智能兴趣组（SIGART），他们发布通讯。AAAI还出版《人工智能》杂志。

PC AI杂志发表有关人工智能技术应用的文章——主要针对决策支持和专家系统。杂志《Engineering Applications of Artificial Intelligence》(EAAI)包括侧重于实时系统的文章。在《IEEE Expert》中有关“智能系统及其应用”的系列文章介绍了世界各地致力于应用研究和工

业技术转化的人工智能实验室。

对不同人工智能论题的总结可查阅《The Encyclopedia of Artificial Intelligence》[Shapiro 1992]、《The Handbook of Artificial Intelligence》中的几卷[Barr & Feigenbaum 1981， Barr & Feigenbaum 1982，Cohen & Feigenbaum 1982，Barr, Cohen & Feigenbaum 1989]以及《Exploring Artificial Intelligence》[Shrobe 1988]。一些人工智能子领域中的重要论文再版在名为“Readings in *X*”的出版物中（*X*代表各种具体领域）。

习题

1.1 给出你自己关于机器的定义。你认为人类是机器吗？不论你的看法如何，运用你的定义和有关人类各种能力的证据证明你的观点。

1.2 你能列举出用蛋白质而不是硅片制造思维机器的实际好处吗？

1.3 假设你是图灵测试中的询问者，想出问*X*或*Y*的五个用于判断它们哪一个是人和哪一个不是人的问题。

1.4 批判地对用图灵测试来判定非人机器是否能思考进行评价，至少提出一种不同观点。

1.5 一些人工智能研究者主张人工智能的目标是建造能“帮助”人们进行智能任务的机器，而不是去“完成”那些任务。不严格地讲，去“帮助”有时被称为“弱人工智能”（*weak AI*），而去“完成”有时被称为“强人工智能”（*strong AI*）。你怎样认为？为什么？