

第 1 章

什么是人工智能

人工智能研究的是如何使机器具有智能行为。为了把人工智能的研究范围定得更加明确，有必要搞清楚什么是智能？智能行为是由什么组成的？同时还必须了解现代意义上的机器一词指的是什么，虽然这比定义智能一词要容易得多。实际上，智能一词并没有什么令人满意的客观的定义，我们在这里虽然也要对这一问题予以讨论，但最终也只能求助于人的直觉想法，再列举出通常在人工智能名下进行的几个研究领域就算完事了。

实际上，机器一词在这里通常指在通用数字计算机上运行的程序。虽然这个词有时也包括一些专用的电子装置，但这些装置实质上也都是计算机之类的东西。这类“机器”的特性和产生于工业革命中的力畜型机器完全不同。不过，近年来出现了一种趋势，那就是人工智能机器日益变得“健壮”起来。人工智能的一个分支、即机器人学就是研究如何使用智能机器来操纵实际世界中的物体。

什么是智能？

对于智能一词的各种定义，心理学家一般都避免作出确

切的回答。有人说过，唯一较中肯的定义是：不管什么样的智力只要能用智力测验测量，都可以算是智能。有些人工智能研究者就接受这种定义，并写出了能解决智力测验中那类问题的程序。

人们显然希望有一个更一般的定义，因而智能有时就被定义为在新情况下能作出恰当反应的能力。某些用来测验动物智力的实验就是按照这一定义设计的。例如，把一只猴子放到一间屋子里，这屋子里挂着一只香蕉，其高度使猴子够不着。再在屋子里放一只箱子，如把箱子放在香蕉底下，猴子站上去就能够得着香蕉。如果这只猴子把箱子拖到香蕉底下，然后站上去取下香蕉，就说明这猴子具有智能。人们还设计了种种更复杂的情况，有时需要猴子用棍子去捞取食品，有时需要它把两根棍子接起来才能从远处捞到吃的。

上述测验中所显示的那种“智能”很接近于我们对智能一词的直觉理解，但这种定义远不够客观。在这里，‘要想恰当地确定一种情况是否算“新的情况”，必须依靠主观的判断。我们让袖珍计算器作两个数相乘的运算。如果这两个数很大而且很陌生，如 π 或一英里的英寸数，该计算器以前可能从未作过这两个数相乘的运算，那末，对于袖珍计算器来说，它所面对的这种情况就可以算是一种“新情况”。但是，一般我们并不会因为它做对了这一运算而认为它有“智能”。这种情况并不象猴子垫起箱子取香蕉那种情况具有质的创新性，但是要用客观的词语将此两种情况区别开来却是困难的。

要想定义“智能”一词，就同人们在试图回答“机器能思维吗？”这一问题时想定义“思维”一词一样困难。人工智能的整个研究课题可以说就是要回答这一问题。今天大多数人不会将袖珍计算器看成是有“智能”或会“思维”的，这件事本身就表明了很重要的一点：制造“智能”机器或“思维”机器的任何努力最终必定会自行归于失败。机器确实可能显示出某些智能行为，但人们会说：“这不是我们所说的思维——我们知道机器怎样做到这一点。”人工智能除非能重复人类的全部智能（这个目标天晓得哪一天才能达到），否则必定会归于失败，就象一个人追逐彩虹一样，他永远不可能追上，虽然他在追逐的过程中有可能发现其他的珍宝。我们上几代的祖先会很容易地把袖珍计算器看成是有“智能”的。（多半看成是有魔力的智能）。但在今天，一部机器要能称得上有智能，就得会做更多的事。

因此，在人工智能这个研究领域，人们试图使机器去做它当前还难于做到的事情，特别是那些本来应该是由人来做的事情。上述人工智能的定义中有“当前难于做到”几个字，这表明人工智能是一个随时间而发展的概念，正如“思维机器”，是由什么构成的这一问题随工艺水平的提高而发展一样。由于人工智能追求这一永远也达不到的目标，由于它吸引了许多能人学者从事这方面的研究工作，因此成了计算机科学中的前沿学科，这一点，在第十三章还会详细谈到。

数学家和计算机科学家 A·图灵（*Turing*）^[1]曾经作过一个很有趣的尝试，藉以表明一架什么样的机器可以算是“思

维机器”。这种尝试是通过所谓的“问答游戏”(参见本丛书的第一本书, F. H. 乔治(George) 著《控制论的哲学基础》一书)进行的。图灵肯定参加过一些非常乏味的晚会,而他的想法起源于晚会上的一种游戏,这游戏要求某些客人悄悄藏到另一间房间里去。然后请留下的人向这些藏起来的人提问题,并要他们根据得到的回答来判定跟他对话的是一位先生还是一位女士。当然,问答必须是间接的,必须有一个中间人把问答写在纸上,或者来回传话(如果有电传打字机,也可用它来作联络工具)。

图灵由此想到,同样可以通过与一架据称有智能的机器作问答来测试这架机器是否真的有智能。如果提问者无法判定与之交谈的是一架机器还是一个人,而实际上是一架机器在与之交谈,那就可以认为这架机器是有了“智能”。

现在的计算机程序可以说已通过了某些简单的图灵测试。某些博弈程序(见第五章和第六章)作出的回答与人作出的回答简直没有什么差别,使得与之交谈的人确信他们是在与一个藏起来的人作问答。有一个名称为 **ELIZA** 的著名的计算机程序,它能扮演精神病医生的角色(见第九章)。“求诊”于这一“精神病医生”的病人中常有人确信这是一位医生,而不是一个程序。

但是,这些程序并不是图灵原来设想的那样,去模拟能随时就各种话题发表见解的具有整体功能的人。博弈程序显然只能完成人的极有限的一部分智能行为。**ELIZA** 程序的功能也很有限,它只是模拟一个医生如何避免向病人透露信

息而引起病人的“预感”，在这方面它倒象个真正的精神病医生。

计算机程序能有效地模拟人的极有限的一部分行为。但人有时会有意不表示出智能来，这就使得图灵测试难以得出明确的结果。图灵测试的另一个不能令人完全满意的地方是它强调智能的模拟，而不是强调取得某些有用的或可能有用的结果（请参见第十五章的最末一段话）。

人工智能是可能的吗？

读者现在也许同意，人工智能确是一个重要的研究领域——既然它确是计算机科学的一门前沿学科，当然就是一个重要的研究领域——但还有两个根本的问题没有得到回答。这两个问题是：

(a) 归在人工智能名下的各种技术真的模拟了大部分我们直觉上理解的那种智能了吗？换言之，人工智能一词是否用的正当？

(b) 人的智能中有没有机器永远也模拟不了的方面？

当然，问题 (a) 与上面提出的问题“机器能思维吗？”很相象。与那个问题一样，问题 (a) 也是很不明晰的，其意义视“智能”一词的直觉含义以及其他含义不清的术语而定。由于问题本身提得不明确，在考虑这一问题时必须防止上述的倾向，不能仅仅因为某种操作是由机器作出的，就否认那是一种智能的表现。

对于具体的任务，比方说下棋，是可能对人的性能和机器的性能作出比较的。有人（其中特别是 *H. L. 德赖弗斯 (Dreyfus)*）¹⁴ 以无可辩驳的事实证明了人的性能中有许多性能是机器无法模拟的。当然，在本节就作出比较还为时过早，这得等到讲完人工智能研究所取得的成果以后（见第十三、十四章）。

但是，有一个问题在这里讨论一下是颇为有益的，即机器是否具有创新能力的问题。人工智能的起源虽然可以追溯到数字电子计算机（例如下棋机和其他自动机）的出现之前，但现代的人工智能研究却几乎都是以数字计算为基础的。这并不奇怪，因为在某些重要的方面，数字计算机远比其他任何的人造物更接近于大脑。它们是迄今为止最为复杂的人工装置。它们在一定程度上也具备大脑那样的概括性，这使得我们无论是对大脑，还是对机器的能力问题，都很难说清。

另一方面，有一种看法认为，计算机必定是低能的，它们所能做到的，只是按照程序设计师为它们所编制的程序而去操作。当然，程序设计师可以在程序中安排一些“选择点”，使其后的操作序列从两种或多种可能性中进行选择而定，而作何选择则取决于前面一系列计算得出的结果。但是，总的思想并没有因此而改变：计算机只能盲目地按照程序设计师为它预先安排好的操作顺序动作。这样一种低能的机器似乎谈不上会有什么“智能”。所谓“智能”一般指的是“创新能力”（这又是一个无法明确定义的词语）。说什

么受程序控制的装置会有创新能力，这种说法的本身似乎就是自相矛盾的。

对计算机始终就存在着两种根本不同的看法，霍夫施塔特^[2]给计算机起了个怪名字，称它们为“灵巧的蠢货”。在第一台数字计算机诞生之时，报刊上用“巨型大脑”之类的词语来形容它，以致于一些负责的科学家感到不得不出发表一种更清醒的看法，强调机器的低能性。但是，几年后，计算机和程序设计技术获得了进一步的发展，上述科学家中又有许多人对他们的原来观点作了修改。

不可否认，计算机确实是低能的程序执行者，但当计算机和程序变得足够复杂时，其行为虽然在原则上是可以预言的，但在实际上却可能无法预言。因而说机器具有潜在的创新能力，这并非是胡思乱想。如果预言一台机器的行为的最简单的方法莫过于实际上开动它（或开动另一台复杂程度相同的机器）来看看会发生什么情况，那么对于可预言行为的这一限制也就显得不那么紧要了。事实上，人脑也是“可预言的”（除了随机噪声效应——见下文）；如果我们知道大脑的每一个神经元（即神经细胞）的特性、它们的相互作用方式以及某个时刻它们的兴奋状态，那么大脑的未来行为也可以是能够预言的。因此，至少可以这样说，由于数字计算机的行为可以预言，从而认为它必然不能显示出智能和创新能力，这种看法是颇成问题的。

上面“可预言的”一词是在并不那么严格的意义上使用的，其中隐含着这样的假定：机器或大脑的输入信号可以根

据以前的输出作出预言。有几种形式方法可以用来表示在指定意义上可预言的系统。如果这些系统的存贮量是有限的，则可表示为有限自动机；如果存贮量没有限制，则一定与图灵机等价。一个计算机程序，根据定义几乎是一台有限自动机（参见本章附录 1 以及本丛中 **F. H. 乔治** 所著《控制论的哲学基础》一书第十章。乔治教授对机器人和人工智能的哲学问题作了详细的论述，因而本书不拟对此再作详细讨论）。

有限自动机和图灵机必须是确定性系统；它们不受诸如随机噪声之类的不可预言效应的影响。有证据表明，大脑受到随机干扰的影响。神经元的兴奋阈限会发生近似随机的波动。据信，大脑中的神经元只能死亡，不能修复。哪些神经元死亡，哪些不死亡 看来也是随机的【**D. 伯恩斯 (Burns)** 估计，成人的大脑每天有十万个神经元死亡】。人造系统也会受随机噪声的影响，因此我们不能光凭这一点来区别人脑和机器。但是，我们确实可以把人脑和在大多数数字计算机上运用的程序区分开，因为后者是被着意设计成确定性的，即不受随机效应的影响。

某些早期的计算机包含有源自某些不可预测效应（如电阻噪声或天然蜕变现象）的真随机数源，因而避开了确定性操作这一限制，但除非在程序中编入一条特殊的指令来引用随机数，否则操作就是确定性的。

现代的数字计算机不包含真随机数发生源。对于需要产生似随机数序列的计算任务，通常使用计算过程来产生这些

似随机数，我们把此种随机数称为“伪随机数”。这些数字并非真正随机，因为它们总是由一个过程从同一周期性序列中产生出来的。事实上，此过程只是将一张非常长的（周期性的）随机数表作极端压缩的存贮。由于此周期很大，程序运行一次所使用的随机数序列只是总随机数表中的一小部分，因此我们从此周期表的不同地方开始，就能得到不同的随机数字序列。当计算机程序的运用需要确定序列中的第一个随机数时，有时我们可以根据现实时钟所标明的时间来决定起始点。这样做就能很方便地在程序每运行一次时便产生出不同的序列，这和随机数产生自真随机数发生源的情况完全一样。

本章附录 2 说明怎样产生伪随机数，我们从中可看出确定性和不确定性之间联系的实质。实际上，对于所谓的“主流”人工智能研究中的大多数程序来说，随机性（或伪随机性）并不是一个重要的特征。但对摆脱第十四章中所讨论的限制而自身能产生出试探法的程序来说，却是大有用处的。

使用“控制论”而编写的人工智能程序常常需要用到随机数或伪随机数。例如，在塞尔弗里奇的群妖堂中使用的“结合法”和“突变分裂法”（见第七章），在种种感知器中；尤其是在罗伯茨提出的一种感知器中（见第八章），都需要用到随机数或伪随机数。在产生“计算机艺术”（见第十一章）的所有方法中，也要用到它。它是程序设计师武库中的有力武器。

由上可见研究人工智能可有种种不同的方法，我们有必

要讨论一下这些方法之间的关系以及它们与控制论的关系。

控制论和人工智能

乔治教授在他的《控制论的哲学基础》等著作中实际上把控制论与人工智能研究等同起来。这种观点同维纳《控制论》^[3]一书副题中出现的控制论意义是一致的，其副题为“动物和机器中的控制与通讯”。

毫无疑问，这里的“动物”一词十分重要，我们从这一新学科中取得的主要成果是它大大增进了我们对于生物体中以及生物社会中所发生的过程的了解。

在这些过程中，大脑的机制问题向我们提出了最大的挑战（在这一方面，甚至还有许许多多最简单的问题都没有能获得解决）。

在探讨控制论的意义从而探讨它的起源时，应该提到这一学科的另一位伟大的开创者 W. 麦卡洛克。虽然维纳的《控制论》一书给这一学科起了名称并且确立起了它的正式地位，但当年积极从事这一领域研究工作的许多人都认为，如给麦卡洛克也冠以“控制论之父”的美名，他是当之无愧的（维纳和麦卡洛克这两位伟大的思想家在控制论正式诞生之时是朋友，哪些贡献是谁做出的，我们不可能分得很清，几年后他们两人之间不幸发生龃龉，其起因不但在现在，就是在当时也是根本说不清的）。

麦卡洛 ^[4]把控制论的发展归因于他试图回答他本人提

出来的一个问题：什么样的数可为人所了解？什么样的人 would 了解一个数？这一问题所包含的意义显然属于人工智能的研究范围，它使我们去探讨认识问题，即外部世界在神经系统或机器内部的反映问题。麦卡洛克坚持认为，认识论应该是“实验的认识论”；他寻求可用神经系统的实验表示出来的特性（一直进行到单元神经元的水平上）来对事物作出解释。

我们看到，乔治教授将控制论与人工智能研究等同起来的这种做法基本上是正确的，但是我们感到困惑不解的是，这两个领域中的许多研究者并不把这两者看作是一码事。大多数从事“主流”人工智能研究的人否认他们与控制论有什么关系，而研究控制论的人则宽宏大量地把人工智能包括在他们的研究范围之内。从本丛书的选题上也可显然看出人们并不把这两个研究领域看作是一回事，如果人工智能就是控制论的话，那么就用不着专门写一本书来论述人工智能了。

前面提到，人工智能的研究有种种不同的方法。由于这些方法可分为两大类，因此倒不如说人工智能的研究出现了分化为二的现象。现在通常所理解的人工智能（本书中称为“主流”人工智能）的研究工作根本打算模拟神经系统中所发生的现象。为研究人工智能而编制出来的大多数计算机程序（这一工作统统是在数字计算机上进行的）与人由经验而学习到的功能并无关系。如果说它们模拟了人的心理活动，也只能说它们是在高水平上模拟人的心理活动。当然，这并不是要否认它们在帮助阐明大脑机制上所具有的潜在价

值。对于大脑这样一个复杂的系统，必须从许许多多不同的角度加以探讨，而且，也许主要得从高水平的观点出发加以探讨。

在控制论发展的早期，一直到 1958 年左右，其研究方法的二分现象一直是不明显的，人工智能（指“主流”人工智能）研究者认为，人工智能研究就是从 1958 年左右所报道的工作开始的。M. 明斯基^[5] (*Minsky*) 和 J. 麦卡锡^[6] (*McCarthy*) 这两人为其后的人工智能研究指明了方向，麦卡锡是程序设计语言 LISP 的设计人（参见第十三章），明斯基当时与麻省理工学院有联系，在那儿他受到了麦卡洛克的影响，因而他能够对研究方法的二分现象持超然态度。那时在进展之中的重要的研究工作还包括纽厄尔 (*Newell*)、肖 (*Shaw*) 和西蒙 (*Simon*) 所作的通用解题程序研制工作（见第三章）以及萨缪尔 (*Samuel*)（见第六章）和塞尔弗里奇 (*Selfridge*)（见第七章）等人所做的工作。

从斯莱格尔^[7] (*Slagle*) 所著的一本导论性教材《人工智能：探试性程序设计法》的标题中我们可以看到研究方法的二分现象。实际上，他关心的是我们这里所谓的“主流”方法，并且想用一个好的简短的术语来称呼“主流”方法，结果他用了探试法一词，这个词很恰当，因为一切采用“主流”方法的研究工作都使用了探试法。

费根鲍姆和费尔德曼^[8] (*Feigenbaum and Feldman*) 在他们合著的一本著名著作《计算机和思维》的序言中也认识到并讨论了这种二分现象。这本书收集了当时（1963年）

人工智能领域最重要的论文，以论文集的形式出版。他们在谈到“主流”方法或“试探程序”方法时用的是“认知模型”这一描述性不很强的术语，他们并把它与“神经控制论”或“自组织系统”方法作了对比，从而作了如下的评论：

“神经控制论用来解决智能机器设计问题的途径是，假定存在着排列成随机网络或有序网络的大量的极简单的信息处理单元，以及某些促进或者抑制它们活动的过程。认知模型的建立者则从更为宏观的角度上来考虑这一问题，他们使用高度复杂的信息处理机制作为设计智能机器的基础。他们相信，要使机器具有智能，不“从头做起”是很难获得成功的。因而他们在系统中引入尽可能复杂的信息处理机制，只要这些信息处理机制他们理解得了，并能（用他们的程序设计方法）传达给计算机。”

费根鲍姆和费尔德曼进而将各种方法取得的成果作了比较并作结论说，“认知模型”已取得了较其他方法远为巨大的进展。事实上，他们声称在“神经控制论”方面的进展（在1963年）“微乎其微”。他们在谈到由简单的信息处理单元构成的网络时，提到了许许多多为产生“自组织系统”而进行的工作，这些工作试图通过与环境相互作用的学习过程获得智能行为。这些研究工作的意图是模拟神经系统的自组织能力，系统的构成单元通常是根据神经细胞的已知特性来设计的。

不用说，这两种方法之间的争论今天还在继续。本书大部分篇幅将用来叙述在“探试程序设计”或“认知模型”名下所进行的研究工作，因为现在与1963年时的情况一样，在这方面有非常丰富的成果可予以报道。但是，也有人提出异议，现在需要的是另一种方法；由于“探试程序设计”方法的特性，对于不同的任务引入了不同的技术，因而成了某些人的攻击对象。当然，对于这些情况的全面讨论，最好在讲完了人工智能已取得的实际成果之后，放到最后几章中去进行。

同时，虽然也许为时迟了一点，还是应该讨论一下人工智能的目的。

人工智能的目的

使机器的行为尽可能地表现出智能（且不管这句话是什么意思）本身便是一项十分迷人的工作，这一领域中的许多工作者很可能从未想到要探究自己的研究动机。而许多其他的研究领域中的研究人员的动机，要末是“解释”某种现象，要末是“利用”某种现象。

有些人公开宣称，他们研究人工智能的目的是通过在计算机程序上模拟人的思维过程来阐明这些过程。有些人甚至表达过这样的意思：如果一位心理学家在自己的工作中不构造这种模型，他就不是一个称职的心理学家。纽厄尔，肖和西蒙的通用解题程序研制工作就是模拟被试的解题行为。被

试一边解题，一边尽可能地把思路讲出来，录入录音机中。通用解题程序就是根据对这些被记录下来的“原话”的内容进行模拟而产生出来的。当然，模拟的准确性决定于解题过程易于反省的程度以及解题过程用语言表达的难易程度。关于这一问题，以后再谈。

有的程序即使不是直接模拟实验中观察到的人的行为，但只要具备所谓的“智能”，也可以说是对人的行为在某种水平上的模拟，从而有可能增进我们对天然智能的了解。

有些人工智能程序所使用的原理与一个人为获得同样结果所使用的方法显然不相一致。例如，大家一般都了解，现有的获得成功的下棋程序并没有模拟棋手所使用的分析方法（见第五章）。但是，即使是对于产生这种程序所需要的解题方法所作的探索，也可能间接地增进我们对于人所使用的方法的了解。这种探索很可能使我们对于问题的实质有所了解，并对某些可能是棋手使用的方法提出若干假设。（但是进行这类推理要特别小心，因为对问题的规模与实质所作的任何估计都肯定受到我们现有方法的限制）；而在自然进化过程中产生出来的种种诀窍中，还有一些显然是现今的程序设计师并没掌握的）。

研究人员的另一类动机是“利用”某些现象。机器的“智能”越强，对使用者就越有用（当然得假定它们服从于人的旨意）。近年来（即在写本书的1981年）由于经济衰退，研究经费被削减，人工智能研究不得不集中于短期中能见到经济效益的项目的研究。因此近年来特别注重于机器

人和专家系统的研制，前者有巨大的经济潜力，后者可应用于医疗诊断及其它方面。此外，各种形式的模式识别也有立时就能获得应用的实际价值。

人工智能的研究课题

在本章及本书目录中我们已经指出了人工智能的各个研究课题，其中主要有：

定理证明：这一课题与其他的数学自动化领域以及其他领域（如机器人学）中的解题方法有一定的交叉。

博弈：主要指下棋。

模式识别：这可以是视觉模式或听觉模式的识别，也可以是其他形式的模式（或混合模式）的识别。医疗诊断和天气预报就并非是只与某一特殊形式的模式相联系的模式识别。在新近的许多研究工作中，重点放在景物分析上，而不是放在孤立物体（例如字母）的识别上。这对于机器人学来说是很重要的。

自然语言的使用：在此研究领域，问答系统和机器翻译受到了很大的重视。威诺格雷特 (*Winograd*) 新近所作的研究工作把自然语言和机器人学联系起来。

机器人学：它有直接的实用价值。

专家系统：这些系统包含了一个专家所具有的大量知识和技能，在医疗诊断以及其它方面有很大的应用价值。

知识工程：知识工程本身还没有形成一个明确的研究

领域，但是对于在模式识别、机器人学和专家系统中应该怎样组织各种不同种类的知识这个问题，它提供了一种特殊的观察角度。

附录 1.1. 有限自动机

所谓自动机可以是有限的，即它能存在于任一有限状态集合中。如果还存在着输入信号的有限集合和输出的有限集合，且此自动机是确定性的，那末对于每一种状态，列出每一输入产生的结果，我们就可把这一自动机表示出来。输入所产生的效果是双重的——它可能产生一输出，同时也能引起自动机状态的改变。

因此，有限自动机 [9] 由五个部分组成。前三个部分为状态集合表、输入表和输出表。第四部分为状态 \times 输入到状态的笛卡尔映射，第五部分为状态 \times 输入到输出的笛卡尔映射。

设一自动机所接受的输入为一个单个的二进数字 (0 或 1) 且接连接接收三个数字 1 才输出一个数字 1。在输出一个数字 1 后，输入的计数回复到 0，要再产生一个输出必须再接连接收到三个数字 1。这样一台自动机一定有四种状态，它们分别与 0 个、1 个、2 个、3 个数字 1 的输入相对应。

所谓笛卡尔映射可用一张表来表示，每一对状态和输入的可能组合对应着一个确定的项。表示上述计数自动机下一状态的变换为：