

人工智能的极限

计算机 不能做什么



休伯特·德雷福斯著

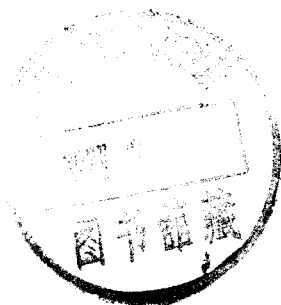
7582
863

休伯特·德雷福斯著

计算机不能做什么

人工智能的极限

宁春岩译 马希文校



生活·讀書·新知 三联书店

8710131

责任编辑：朱志焱

封面设计：王师颜

Hubert L. Dreyfus

WHAT COMPUTERS CAN'T DO

revised edition

The Limits of Artificial Intelligence

Harper & Row, Publishers, 1979

译自美国哈泼·罗出版社 1979 年修订版

计算机不能做什么

——人工智能的极限

〔美〕休伯特·德雷福斯著

宁春岩译 马希文校

生活·读书·新知三联书店出版

北京朝阳门内大街166号

新华书店发行

北京新华印刷厂印刷

850×1168毫米32开本 11.5印张 275,000字

1986年10月第1版 1986年10月北京第1次印刷

印数 00,001—10,000

书号 2002·271 定价 2.25元

校者的话——代中译本序

毫无疑问，通用数字计算机的发明是人类历史上最重要的事件之一。计算机的广泛应用正在深刻地改造着生产力的面貌，这又不可避免地会对生产关系带来重大的影响。如果有人认为，由此将会使资本主义生产关系的物质基础即社会化的大生产发生动摇，并为马克思和恩格斯在《共产党宣言》中描述的那种更先进的共产主义生产关系准备了物质基础，那么我们不但不必惊奇，而且应该认真研究一下这是否正是社会发展的大势，如果是这样，那么我们应该怎样去迎接它、推动它。

无论计算机对人类的贡献能不能与当年的蒸汽机相比，它在我国当前的改革事业中总占有一个重要的地位，成为上层建筑改革的物质基础，对改革起推动作用。作为共产主义初级阶段的社会主义制度，其优越性归根结底应表现为能适应社会发展的客观规律，逐步向共产主义的高级阶段过渡，以实现“交往方式的生产”。因此，重视计算机在新技术革命中的地位、注意计算机应用对改革的影响，不但是必需的，而且是必然的。总之，某种形式的计算机热一定会在我国出现，事实也正是如此。

然而，推广计算机时，应采取科学的态度，特别是在对计算机

的社会作用的评价方面。

比如说，七十年代末期，人们过分强调了计算机在生产自动化方面的作用，甚至把计算机应用的目的片面地解释为节约劳动力。这就使许多人担心这会带来劳动力过剩的问题，继而对之产生疑问，终致使我国计算机应用的日程表向后推迟了若干年。

如果我们认真从这件事中吸取了教益，那就应该慎重地对待当前在宣传中越来越普遍使用的“电脑”这一绰号——这如果不算是“港台风”，也至少使我们想起四十年代有人把摩托车称为“电驴”。不应只从商品心理学的角度谈论这个问题，因为它暗示了计算机可以取代人脑这一重大的哲学命题。实际上，许多学者已经明明白白地把这个命题表述出来了。如果漫不经心地把没有经过科学论证过的命题应用到社会生活中去，我们很可能会再一次付出代价。

其实，主张计算机能代替人脑的，不仅我国学术界有，国际学术界更是大有人在。特别值得注意的是他们当中包括大部分的知名学者。这就使许多年轻人或外行人更容易轻信。

在这种形势下，本书以少数派的姿态出现，力排众议，对这种主张进行了详细的分析和批判，引起了西方学术界的注意。我们把这本书介绍给我国的读者，也是为了引起大家的关心和讨论。

我们对于人工智能界中的所谓乐观派持批判态度。这是我们与作者的共同之处。但是，从什么立场出发进行这种批判，则有很大差别。此外，作者在本书中采用的论战方式，加上他那刻薄、讥讽的语言（以及翻译时的种种技术性困难），可能使读者读本书时感到吃力。基于以上两方面的理由，我们写了这篇序言。本书主要是从哲学的角度立论的，我们则偏重技术性的讨论。从这方面来说，也许又为读者提供了一些补充材料。

二

计算机能不能代替人脑？这是一个十分含糊的问题。至少可以有以下几种很不同的解释：

(1) 计算机能不能完成一些迄今为止主要是靠人的大脑的活动完成的工作？

(2) 计算机能不能完成一切这种工作？

(3) 计算机能不能象大脑一样地完成这种工作？

对于(1)，回答是肯定的，这已由当前的社会实践所证实了。

其实，自计算机发明之日起，它所做的一切工作一直是那些主要靠人的大脑完成的工作，例如科技计算，文件编辑，信息管理等。当前，计算机的应用领域仍在不断扩大。随着计算机技术的进展以及使用方法越来越丰富多样、灵活巧妙，计算机的应用正把它的触角伸向那些通常以为很难使用计算机的领域，如专家系统、机器翻译、模式识别、定理证明、问题求解、自然语言理解等等。这些是计算机应用的前缘，它们构成了人工智能学科的主要内容。这些方面的成就，本书中分为四个时期作了概述。

这方面的任何进展，对于熟悉计算机原理，懂得程序设计的人来说，都不会感到神秘。如果有时感到意外的话，那大多是因为对应用领域不熟悉，或充分意识到了技术上的困难。计算机是进行信息处理的，其中的“信息”就是指有限种符号的有限长序列这种形式的信息，而“处理”的过程就是按预先编好的程序对这种序列做有穷的变换，以得到一组新的符号做为结果。这就是所谓“计算”（包括数值的和非数值的计算）。但是许多具体问题怎样用这样的办法来求解并不是一目了然的，往往需要高超的技巧。例如说，怎样编制一个能在使用过程中自行改进的程序就是这样的例

子。当本书中介绍的能自行改进的跳棋程序战胜了它的设计者的时候，人们感到意外，是自然的。

系统地发展这一类的技巧，提高计算机应用的灵巧性，以便扩大计算机的应用，不但是可能的，而且是十分有意义的。计算机的社会意义正是产生于它应用的深度和广度可以不断得到发展，取得许多重要的成就。除了可计算性（包括理论的和实际的可计算性）之外，几乎再也找不到一个合理的说法来为计算机的应用划定一个技术性的界限了。

由于这种情况，只要稍不谨慎，就会陷入幻觉，以为计算机的应用并无界限，也就是说，对前面的问题（2）做出肯定的回答。这就是书中说的 AI 学派的论点。然而这是一种没有根据的外推，其中包含着不合理的逻辑跳跃。下面我们就来仔细地分析一下这个问题。

三

要计算机去解决某种问题，有三个基本的前提：

第一，必须把问题形式化。计算机，至少在它较低的层面上，只能进行有限符号集上有限长符号串的决定性的形式变换。人们在使用计算机时，常常从观念上以及实现手段上加上一些较高的层面。因此上面的要求可以稍稍放宽，比如可以不必事先假定符号集是有限的，可以认为符号串是有结构的等等。但是无论如何放宽，计算机总是在做符号处理。因此，任何问题，要交给计算机去解，必须先建立一个形式系统，规定所用的符号、规定符号联结成合法符号串的规则（语法）、以及合法符号串如何表示问题领域中的意义（解释）。然后，建立一些规则，说明对这些符号串可以进行一些什么样的处理。于是，要解的问题就可以用符号串表示出

来，怎样算是一个解，也可以表现为对符号串的一些要求或条件。这样一来，计算机求解的过程就是从表示问题的符号串出发，按规则进行加工，直到得出符合要求的符号串为止。

这一整套办法简称为形式化。计算机的各种应用无不是靠了这种办法实现的。形式化方法的实用范围比人们最初预料的要广泛得多。

然而，按照AI学派的主张，计算机应能解决人脑能够解决的一切问题。那么，它也应能解决如何把某种问题形式化的问题。但这样一来，就出现了一系列的困难。

首先，形式化是从非形式化的领域向形式化的领域的转变。如要计算机来完成这一转变，就得把这个转变形式化，这又包括了如何表示转变的起点——一个非形化的领域。这就造成了一种回归现象(甚至是谰论)。

要想避免这种回归，必须假定我们有一种包罗万象的先验的形式化系统。但形式化系统属于理性认识的范畴，不是先验地存在的。换言之，要计算机能代替人脑的一切工作，就要预先把我们今天尚未认识而明天可能认识的领域预先形式化，这是做不到的。

即使做到了这一点，由于客观世界的无限丰富性，我们也会遇到一个无限多的符号、无限多的规则的形式系统。这与计算机的资源(存贮空间等)的有限性是矛盾的。

当然，人们可以反驳说，我们可以把问题领域逐个形式化。但这又引起了一个新的问题，即计算机遇到某个问题时，如何判断该问题所属的领域的问题。这个新的问题是比原来的问题更高一个层面上的问题。这也导致无穷的回归。作者多次论及如何根据局势或环境判断有关性和重要性，这个问题就与此有关。

总之，指望把人脑能做的一切交给计算机去完成，首先遇到的也是最基本的困难就是形式化的困难。

第二，计算机要解决已形式化的问题，问题还必须是可计算的，即一定要有算法。这个问题作者没有明晰地论述，但他所说的许多问题都与此有关。完全的定理证明系统之不可能，可以从理论上予以证明；判定刁藩都方程是否有解的算法，也可以从理论上证明其不存在。这些都为计算机代替人脑设下了陷阱。

另一方面，存在着某个算法和找出这个算法是两回事，前者是客观的，后者是人脑的功能。要想用计算机代替人脑，就要找出一种算法来代替人脑寻找算法，这又遇到了回归。

有人可能争辩说，我们有搜索法，这是一种万能的算法。这其实是一种误会。搜索法在涉及无穷集合的问题中无法施展。比如说，给定一个自然数之后，我们可以用搜索的办法来判断它能否分解为两个奇素数之和，因为我们只用检查它的每种分解式中的两个数是否奇素数就行了，而这种分解式是有穷多的。要想用同样的搜索法来判断它能否分解为两个奇素数之差却办不到，因为这种分解式是无穷多的。其实，如果给定的数是奇数，一点点关于奇偶性的考虑便可以证明它不能分解为两个奇素数之和或差。这个证明当然也可以用搜索法发现，但却要在一个稍许不同的形式系统中去做。如果给定的数是偶数，这已是“哥特巴赫猜想”这样的难题，要找寻这样一个证明，则要有更复杂的形式系统。从符号串的变换角度来观察这个问题，它们是分属于不同的形式系统的问题，虽然从解释上来看，它们是一致的，这就是在不同的层面上看待同一问题。一个问题（比如证明一个命题），只有当它在某个形式系统中可解的时候，才能指望着用搜索的办法求出解来，在涉及无穷集合的问题中，如果问题无解，搜索过程并不能发现它无解。因此，当我们为了解答问题而不得不从一个层面即形式系统向另一个层面（也是形式系统）过渡时（且不说这个形式系统怎样做出的问题），我们无法判断应在哪一个层面上停止下来。这样，搜索法

也就失灵了。

第三,要用计算机实现一个算法以解决某种问题,问题必须有合理的复杂度。这常被说成是避免指数爆炸。是否会发生指数爆炸的情况是问题本身的性质,不是任何巧妙的技术可以绕过的。

人工智能目前面临的多数难题都涉及指数爆炸。用搜索法求解问题更是如此。多年来,一些人工智能学者对这个问题佯装不知,这实在荒谬得令人不解。其结果是许多巧妙的设想,在一定的规模内似乎灵验,一旦问题规模稍稍扩大,计算机就无法胜任。这就是作者说的“初期的成功与随之而来的停步不前”的重要原因。本书对这个问题说得不多,但王浩在《数理逻辑通俗讲话》中对AI学派的批判主要就集中在这一方面。

由于计算机解题必须具备以上三方面的条件,所以要计算机完成人脑能做的一切工作是毫无希望的。

四

既然问题(2)已有了否定的答案,(3)的答案似乎也就是否定的了。其实并不这样简单。

计算机多次展示了这样的能力,即虽然程序是人设计的,但计算的结果却是出乎意料的,甚至比人原来设想的更加高明。我这里指的是象下棋程序战胜它的设计者、定理证明系统找出某个熟知定理的新颖的(至少对它的设计者来说是不曾想到的)证明这一类的情况。这样,书中称为认知模拟学派或CS学派的那些学者,就产生了一种希望,即不必先把问题形式化,只要我们给计算机配上一套能模拟大脑活动的程序,就可以造出思维。对这种观点进行反驳,上文的讨论就不够了。

从逻辑上看,当我们说到计算机代替人脑时,的确并不意味着

计算机能解决一切问题。在任何一个具体环境下的具体的人，都不是万能的。因此也没有理由要求计算机是万能的。如果我们谈论的不是代替人脑的一切潜在能力，而只是代替某一具体人脑的能力时，问题的性质当然是极不相同的。

为此，书中分析了人类的四种“信息处理”活动，指出这些是任何正常人脑都具有的能力，而计算机却无法做到这些。我们则打算从另一角度来讨论这个问题。

CS学派和AI学派有一种根本的区别：AI 必须把客观世界做成形式模型，而CS则只用把大脑做成形式模型就够了。

大脑是物理世界的东西，它的活动首先是物质的运动。这就是大脑的物理的、化学的、生物学的活动。用分子生物学、遗传工程、人造蛋白质等手段来造出一个大脑，这也许会成功，但总不能算是计算机代替了人脑。要从物质运动的层面上模拟大脑，就要先写出大脑活动的数学模型，并在计算机上模拟它。这就要求搞清大脑的结构和机制，就会出现大量的细节。比如说某个脑细胞的构造吧。它有一定的外部结构（如形状，重量），细胞学层面上的结构（如细胞核、染色体），其中每个组成部分又有分子生物学层面上的结构（如分子的空间构形），每个原子又有基本粒子层面上的结构，如此等等，以至无穷。到底哪些东西是大脑结构中实质性的东西呢？又如脑细胞之间的联系，除了宏观的电、磁、力等方面的联系之外，还有分子的交换以致基本粒子的交换等等，到底哪些是实质性的呢？如果对于这些没有清楚的了解，如何去模拟大脑的活动呢？

当然，多数 CS 学派的学者并不想在这个层面上模拟大脑，大脑的思维功能也不可能从这个层面来描述，正如计算机的功能不能从分子运动的层面来描述一样。他们设想在物质运动的层面之上，还有一个信息处理的层面，再上面又有认知的层面等等，然后

才到达思维的层面。这当然只是一种假说，无论从理论上或实验上都远未得到证实。恐怕还要许多代人的努力才能有比较清晰的认识。可能还会添加许多层面，也许这些层面并不象是一条直链上的许多中间环节，而是一个网络中的不同结点。总之，认为信息处理是从大脑的生理活动到达思维的必经之路是没有根据的。尤其是本书中提到大脑并不象三十年前设想的那样可以看成一套复杂的开关网络。这种实验证据十分值得注意，因为它可能改变人们对于信息处理在大脑活动中作用的认识。近年来，人们开始注意形象思维（以及顿悟）的特殊规律，看来，用信息处理对此作出解释的可能性是大有疑问的。

计算机是信息处理的机器，要用它来模拟大脑，最自然的莫过于在信息处理的层面上进行模拟。但是上面的讨论指出，这样做将引起双重的困难：

（1）大脑的功能是否都可以通过信息处理活动来描述？

（2）从大脑的信息处理层面到达认知的层面是否还有许多别的中间环节？

书中讨论生物学和心理学的假设时，详细阐明了作者与此有关的论点。

另一方面，大脑的功能根本不是只由其生理的机能决定性地发展出来的，而需要与周围环境发生联系。孤立地模拟大脑的机制，不把它放到与外界相联系的环境之中，当然不能产生思想。因此又要模拟或实现感觉器官、运动器官的功能，这样甚至又无法避免外部世界的形式化的问题。作者把这些表述为“躯体的作用”，内行人可能觉得可笑。但是问题不在于躯体的形状如何，而在于它的功能，脱离了这种功能，大脑（不管是人脑还是电脑）不只会遇到能不能存在的问题，而且根本就不可能产生思想。

五

我们谈到了计算机在人类历史上的独特地位，因为它的发明使我们能在一定意义上有了一种人工的制品、一种机器或工具，可以用来“延长”我们的大脑。但我们又花了较多的篇幅来说明，计算机的结构、工作方式都与大脑迥异，其功能也远不能达到大脑做到的一切。认为已经揭开了或至少可以猜测到大脑与思维的奥秘，这是毫无根据的。因此，我们不应止于讨论大脑与计算机之异同，而应关心人工智能的未来。本书末尾对此提出了很值得注意的见解。

计算机不应是也不会是最终的智能机器。人工智能要向前迈进，就不应把自己局限于计算机的应用。这里指的是以图灵机和可计算理论为背景的现代通用数字计算机。应该开创一门新的学科（把它叫思维科学或智能科学都行），研究思维活动的更深入的具体规律，提出新的概念、新的方法和新的机制，比信息处理更为广泛、更为深入地描述思维的某些功能，并把这些与某种（理论的）机器模型相联系以期最终得到工程实现。

从古人发明四则运算到手摇计算机的出现，人类经历了一次从对思维的认识发明智能机器的过程，这是不自觉的，花了一千多年。今天看来，这可算是思维科学的一段史前史。从巴贝吉、图灵到今天的计算机，又经历了一次这样的过程。这一次花了一百多年，思维科学可以说就此起步了。然而对思维的认识是一个不断前进的无穷过程。每推进一步，就更加接近于思维的本来面目。盲目的乐观固然不对，盲目的悲观也没有根据。因为两者都使我们停止不前，而我们不能总停留在一个水平上。

当然，现代计算机的潜力远未完全发掘出来。即使就形式系

统中的信息处理这个层面而论，我们对思维的认识也尚未完成。认知心理学、认知逻辑、理论语言学等多种学科都在这个领域中努力向前推进。人工智能在这种形势下应对信息处理，计算机的能力做许多具体的研究，求得更加明晰、准确的认识。从应用上来看，谈论人脑与计算机的彼此替代未免空泛、消极，不如研究使两者取长补短的人机共生系统。这样做，不只有实用意义，而且对于我们对思维的认识、对信息处理在思维中的地位的认识将提供许多有启发性的实验资料。

本书出自哲学家的手笔。对人工智能的未来提出了一些原则性的见解。其中某些论点也许并不准确。但提出计算机的能力是有限度的这一命题，并进一步指出应当研究新的智能机器，则是我们可以从中学到的极有价值的东西。我们在积极扩大计算机应用的同时，应该想到这些，这样才能有见识地去研究思维的其他侧面和更高层面，推动思维科学的前进。新型的智能机器必将出现，那将不是现代的通用数字计算机，虽然它也不会是人类的最终发明，但却更有资格称为“电脑”——如果不是“磁脑”或“光子脑”的话。

马希文

一九八四年九月

目 录

序	1
致谢	4
修订版序言	5
前言	75

第一部分 人工智能研究十年(1957—1967)

一、第一阶段(1957—1962): 认知模拟	99
1. 对语言翻译、问题求解和模式识别研究 工作的分析	99
2. 未能取得预期成果中所包含的意义	108
结论	138
二、第二阶段(1962—1967): 语义信息加工	140
1. 语义信息加工程序分析	142
2. 现有困难的含义	156
结论	158

第二部分 坚定的乐观主义后面的假想

序言	165
三、生物学假想	168
四、心理学假想	172
1. 心理学假想的经验性证据: 对认知模拟的	

DL83/F1

科学方法论的批判	178
2. 心理学假想的先验性论断	183
结论	195
五、认识论假想	197
1. 从物理学成功中所得出的错误论断	199
2. 从现代语言学成功中所得出的错误论断	205
结论	210
六、本体论假想	214
结论	233

第三部分 传统假想的更换

序言	239
七、躯体在智能行为中的作用	243
八、局势：不依赖规则的有序行为	264
九、作为人类需要之函数的局势	281
结论	289

结论 人工智能的范围和极限

十、人工智能的极限	293
十一、人工智能的未来	299
注解	313

序

德雷福斯选定要对人工智能做一番剖析。他已在这方面做出了十分引人注目的探索。人工智能这一学科被其实际工作者们看作是获取更有力的、早已渗透在我们社会中的、计算机技术的基础。正象人们在阅读日报时看到的那样，许多人在数字计算机所引起的恐惧和希望之间踌躇徘徊。他们感觉这些数字计算机几乎完全不可思议，因而也就无法判断它们的意义。但是，正如科学需要人们的支持那样，人们也需要对科学的批判分析。

德雷福斯对于我们来说是一位富有批判精神的旁观者，一位热心于探究和分析知识基础问题的专业哲学家。他勇敢地步入了一个神秘的技术领域。因此，人工智能界决不可把他当作不受欢迎的冒犯者而拒之门外，也不要挖苦的言辞非难他，而应该欢迎他，汲取他正确的见地，小心翼翼地然而又是踏踏实实地、开诚布公地把那些可能有损于他的逻辑的误解及不理解的地方纠正过来。德雷福斯提出了一些重要的和带有根本性的问题。因而，人们可能会期待对于他的批评能以更多的智能做出反应，而不只是高喊几句他从前讲过挖苦之词。对于他所提出的问题需要进行严肃的公开辩论。这些问题带有过多的科学性质，不能推委给哲学家；但又带有过多的哲学性质，又不能推委给科学家。

德雷福斯看到所有的人工智能基础研究进展得都十分缓慢。他把这看作是存在着不可逾越障碍的标志，而不是那种为克服取