

第1章 绪 论

本章我们试图解释为什么我们认为人工智能是一个最值得学习的学科，并试图让学生们对人工智能的定义有一个简单的认识，从而开启人工智能的学习之旅。

1 人工智能基本概念

1.1 人工智能的定义

众所周知，相对于天然河流（如亚马逊河和长江），人类开凿了叫做运河（如苏伊士运河和中国大运河）的人工河流；相对于天然卫星（如地球的卫星——月亮），人类制造了人造卫星；相对于天然纤维（如棉花、蚕丝和羊毛），人类发明了维尼纶和涤纶等人造纤维；相对于天然心脏、天然婴儿、自然受精和自然四肢等，人类创造了人工心脏、试管婴儿、人工授精和假肢等人造物品（artifacts）……2009年7月8日，英国一个科学研究小组宣布首次成功利用人类干细胞培育出成熟精子。这就是人工精子，一种很高级的人工制品。我们要探讨的人工智能（artificial intelligence），又称为机器智能或计算机智能。由此可见，人工智能本质上有别于自然智能，是一种由人工手段模仿的人造智能；至少在可见的未来应当这样理解。

像许多新兴学科一样，人工智能至今尚无统一的定义，要给人工智能下个准确的定义是困难的。人类的自然智能（人类智能）伴随着人类活动时处处存在。人类的许多活动，如下棋、竞技、解答题、猜谜语、进行讨论、编制计划和编写计算机程序，甚至驾驶汽车和骑自行车等，都需要“智能”。如果机器能够执行这种任务，就可以认为机器已具有某种性质的“人工智能”。不同科学或学科背景的学者对人工智能有不同的理解，提出不同的观点，人们称这些观点为符号主义(symbolism)、连接主义(connectionism)和行为主义(actionism)等，或者叫做逻辑学派(logicism)、仿生学派(bionicsism)和生理学派(physiologism)。在1.2节将综述他们的基本观点。

哲学家们对人类思维和非人类思维的研究工作已经进行了两千多年，然而，至今还没有获得满意的解答。下面，我们将结合自己的理解来定义人工智能。

1.1.1 人工智能的定义

定义 1.1 (智能: intelligence). 人的智能是人类理解和学习事物的能力或者说，智能是思考和理解的能力而不是本身做事的能力。

定义 1.2 (智能机器: intelligence machine). 智能机器是一种能够呈现出人类智能行为的机器，而这种智能行为是人类用大脑考虑问题或创造思想。

另一种定义为：智能机器是一种能够在不确定环境中执行各种拟人任务（anthropomorphic tasks）达到预期目标的机器。

定义 1.3 (人工智能学科). 长期以来，人工智能研究者们认为：人工智能（学科）是计算机科学中涉及研究、设计和应用智能机器的一个分支，它的近期主要目标在于研究用机器来模仿和执行人脑的某些智力功能，并开发相关理论和技术。

近年来，许多人工智能和智能系统研究者认为：人工智能（学科）是智能科学(intelligence science)中涉及研究、设计及应用智能机器和智能系统的一个分支，而智能科学是一门与计算机科学并行的学科。

人工智能到底属于计算机科学还是智能科学，可能还需要一段时间的探讨与实践，而实践是检验真理的标准，实践将做出权威的回答。

定义 1.4 (人工智能能力). 人工智能（能力）是智能机器所执行的通常与人类智能有关的智能行为，这些智能行为涉及学习、感知、思考、理解、识别、判断、推理、证明、通信、设计、规划、行动和问题求解等活动。

1950年图灵(Turing)设计和进行的著名实验（后来被称为图灵实验，Turing test），提出并部分回答了“机器能否思维”的问题，也是对人工智能的一个很好注释。

为了让读者对人工智能的定义进行讨论，以更深刻地理解人工智能，下面综述其他几种关于人工智能的定义。

定义 1.5 人工智能是一种使计算机能够思维，使机器具有智力的激动人心的新尝试[Haugeland, 1985]。

定义 1.6 人工智能是那些与人的思维、决策、问题求解和学习等有关活动的自动化[Bellman, 1978]。

定义 1.7 人工智能是用计算模型研究智力行为[Charniak and McDermott, 1985]。

定义 1.8 人工智能是研究那些使理解、推理和行为成为可能的计算[Winston, 1992]。

定义 1.9 人工智能是一种能够执行需要人的智能的创造性机器的技术[Kurzweil, 1990]。

定义 1.10 人工智能研究如何使计算机做事让人过得更好[Rick and Knight, 1991]。

定义 1.11 人工智能是研究和设计具有智能行为的计算机程序，以执行人或动物所具有的智能任务[Dean et al., 2003]。

定义 1.12 人工智能是一门通过计算过程力图理解和模仿智能行为的学科[Schalkoff, 1990]。

定义 1.13 人工智能是计算机科学中与智能行为的自动化有关的一个分支[Luger and Stubblefield, 1997]。

其中，定义 1.5 和定义 1.6 涉及拟人思维；定义 1.7 和定义 1.8 与理性思维有关；定义 1.9~定义 1.11 涉及拟人行为；定义 1.12 和定义 1.13 与理性行为有关。

1.1.2 图灵测试

1950 年，英国数学家 Alan Turing 提出了一个测试方法来确定一个机器能否思考。该方法需要两个人对机器进行测试，其中，一人扮演提问者；另一人作为被测人员。这两个人与机器分别处在 3 个不同的房间，提问者通过打印问题和接收打印问题来与被测人员和被测机器进行通信。提问者可以向被测机器和被测人员提问，但他只知道接受提问的是 A 或 B，而不知道被测者是人还是机器，并试图确定谁是机器、谁是人。这个测试后来被人们命名为“图灵测试”。

在“图灵测试”中，如果人的一方不能区分对方是人还是机器，那么就可以认为那台机器达到了人类智能的水平。要通过这个测试，那么它的目标就是要使得提问者误认为它是人。因此，有时机器要故意伪装一下，如当提问者问“12324 乘 73981 等于多少？”时，机器应等几分钟回答一个有点错误的答案，这样才更显得像人在计算。当然，一台机器要通过图灵测试主要的是它具有的知识总量和具有大量的人的基本常识。

1.2 人工智能的发展历史

不妨按时期来说明国际人工智能的发展过程，尽管这种时期划分方法有时难以严谨，因为许多事情可能跨接不同时期，另外一些事件虽然时间相隔甚远但又可能密切相关。

1.2.1 人工智能起源

无论是计算机科学还是它的分支——人工智能，其最初的理论基础都归功于图灵。图灵是英国数学家、逻辑学家，1912 年生于伦敦少年时就表现出独特的直觉创造能力和对数学的爱好，被称为计算机科学之父，人工智能之父。

1936 年 5 月，图灵向伦敦权威的数学杂志投了一篇论文，题为《论数字计算在决断难题中的应用》。该文于 1937 年在《伦敦数学会文集》第 42 期上发表后，立即引起广泛的注意。图灵在这篇文章中定义了后来被他的导师丘奇称为“图灵机”的计算装置：一条无穷长的纸带，一个读写头在一个控制装置的控制下在纸带上方左移右移，读取纸带上的内容并在纸带上写 0 或 1，并由此有了所谓“丘奇-图灵论题”，也就是说任何计算装置都等价于图灵机。这个论题不是数学定理，但却是整个计算机科学的基础。

有了“图灵机”，我们就很容易把原来是纯逻辑或纯数学的东西（例如递归函数和 λ 演算等）和物理世界联系起来了，函数成了纸带和读写头。如果按照传统哲学的说法：数学唯心，理性；物理唯物，经验，那么计算机科学就是唯物和唯心的桥梁，司马贺把这类东西统统叫作“人工科学”。

1948 年，图灵在英国国家物理实验室的内部报告中区分了“肉体智能”和“无肉体智能”的概念；1950 年，图灵进一步提出关于机器思维的问题，他的论文《计算机和智能》，引起了广泛的注意和深远的影响。

1950 年 10 月，图灵发表论文《机器能思考吗？》这一划时代的作品，使图灵赢得了“人工智能之父”的桂冠。

1956 年 8 月，在美国汉诺斯小镇宁静的达特茅斯学院中，约翰·麦卡锡、马文·闵斯基（人工智能与认知学专家）、克劳德·香农（信息论的创始人）、艾伦·纽厄尔（计算机科学家）、赫伯特·西蒙（诺贝尔经济学奖得主）等科学家聚集在一起，讨论着一个完全不食人间烟火的主题：用机器来模仿人类学习以及其他方面的智能。会议足足开了两个月的时间，虽然大家没有达成普遍的共识，但是却为会议讨论的内容起了一个名字：人工智能。因此，1956 年也就成为了人工智能元年。

1.2.2 人工智能发展阶段

1、孕育时期（1956 年前）

人类对智能机器和人工智能的梦想和追求可以追溯到三千多年前。早在我国西周时代（公元前 1066—前 771 年），就流传有关巧匠偃师献给周穆王一个歌舞艺伎的故事。作为第一批自动化动物之一的能够飞翔的木鸟是在公元前 400 年——前 350 年间制成的。在公元前 2 世纪出现的书籍中，描写过一个具有类似机器人角色的机械化剧院，这些人造角色能够在宫廷仪式上进行舞蹈和列队表演。我国东汉时期（公元 25—220 年），张衡发明的指南车是世界上最早的机器人雏形。

我们不打算列举三千多年来人类在追梦智能机器和人工智能道路上的万千遐想、实践和成果，而是跨越三千年转到 20 世纪。时代思潮直接帮助科学家去研究某些现象。1913 年，年仅 19 岁的维纳（Wiener）在他的论文中把数理关系理论简化为类理论，为发展数理逻辑作出贡献，并向机器逻辑迈进一步，与后来图灵（Turing）提出的逻辑机不谋而合。1948 年维纳创立的控制论（cybernetics），对人工智能的早期思潮产生了重要影响，后来成为人工智能行为主义学派。数理逻辑仍然是人工智能研究的一个活跃领域，其部分原因是由于一些逻辑-演绎系统已经在计算机上实现过。不过，即使在计算机出现之前，逻辑推理的数学公式就为人们建立了计算与智能关系的概念。

丘奇（Church）、图灵和其他一些人关于计算本质的思想，提供了形式推理概念与即将发明的计算机之间的联系。在这方面的重要工作是关于计算和符号处理的理论概念。1936 年，年仅 26 岁的图灵创立了自动机理论（后来人们又称为图灵机），提出一个理论计算机模型，为电子计算机设计奠定了基础，促进了人工智能，特别是思维机器的研究。第一批数字计算机（实际上为数字计算器）看来不包含任何真实智能。早在这些机器设计之前，丘奇和图灵就已发现，数字并不是计算的主要方面，它们仅仅是一种解释机器内部状态的方法。被称为人工智能之父的图灵，不仅创造了一个简单、通用的非数字计算模型，而且直接证明了计算机可能以某种被理解为智能的方法工作。

事过 20 年之后，道格拉斯·霍夫施塔特(Douglas Hofstadter)在 1979 年写的《永恒的金带》(An Eternal Golden Braid)一书对这些逻辑和计算的思想以及它们与人工智能的关系给予了透彻而又引人入胜的解释。

麦卡洛克(McCulloch)和皮茨(Pitts)于 1943 年提出的“拟脑机器”(mindlike machine)是世界上第一个神经网络模型（称为 MP 模型），开创了从结构上研究人类大脑的途径“神经网络连接机制”，后来发展为人工智能连接主义学派的代表。

值得一提的是控制论思想对人工智能早期研究的影响。正如艾伦·纽厄尔(Allen Newell)和赫伯特·西蒙(Herbert Simon)在他们的优秀著作《人类问题求解》(Human Problem Solving)的“历史补篇”中指出的那样，20 世纪中叶人工智能的奠基者们在人工智能研究中出现了几股强有力的思潮。维纳、麦卡洛克和其他一些人提出的控制论和自组织系统的概念集中讨论了“局部简单”系统的宏观特性。尤其重要的是，1948 年维纳发表的《控制论——关于动物与机器中的控制与通信的科学》，不但开创了近代控制论，而且为人工智能的控制论学派（即行为主义学派）树立了新的里程碑。控制论影响了许多领域，因为控制论的概念跨接了许多领域，把神经系统的工作原理与信息理论、控制理论、逻辑以及计算联系起来。控制论的这些思想是时代思潮的一部分，而且在许多情况下影响了许多早期和近期人工智能工作者，成为他们的指导思想。

从上述情况可以看出，人工智能开拓者在数理逻辑、计算本质、控制论、信息论、自动机理论、神经网络模型和电子计算机等方面做出的创造性贡献，奠定了人工智能发展的理论基础，孕育了人工智能的

胎儿。人们将很快听到人工智能婴儿呱呱坠地的哭声，看到这个宝贝降临人间的可爱身影！

2、形成时期（1956——1970 年）

到 20 世纪 50 年代，人工智能已躁动于人类科技社会的母胎，即将分娩。1956 年夏季，由年轻的美国数学家和计算机专家麦卡锡（McCarthy）、数学家和神经学家明斯基（Minsky）、IBM 公司信息中心主任（Lochester）以及贝尔实验室信息部数学家和信息学家香农（Shannon）共同发起，邀请 IBM 公司莫尔（More）和塞缪尔（Samuel）、MIT 的塞尔夫里奇（Selfridge）和索罗蒙夫（Solomonff），以及兰德公司和 CMU 的纽厄尔和西蒙共人，在美国的达特茅斯（Dartmouth）大学举办了一次长达两个月的十人研讨会，认真热烈地讨论用机器模拟人类智能的问题。会上，由麦卡锡提议正式使用“人工智能”这一术语。这是人类历史上第一次人工智能研讨会，标志着人工智能学科的诞生，具有十分重要的历史意义。这些从事数学、心理学、信息论、计算机科学和神经学研究的杰出年轻学者，后来绝大多数都成为著名的人工智能专家，为人工智能的发展作出了重要贡献。

最终把这些不同思想连接起来的是由巴贝奇（Babbage）、图灵、冯·诺依曼（von Neumann）和其他一些人所研制的计算机本身。在机器的应用成为可行之后不久，人们就开始试图编写程序以解决智力测验难题、数学定理和其他命题的自动证明，下棋以及把文本从一种语言翻译成另一种语言。这是第一批人工智能程序。对于计算机来说，促使人工智能发展的是什么？是出现在早期设计中的许多与人工智能有关的计算概念，包括存储器和处理器的概念、系统和控制的概念，以及语言的程序级别的概念。不过，引起新学科出现的新机器的唯一特征是这些机器的复杂性，它促进了对描述复杂过程方法的新的更直接的研究（采用复杂的数据结构和具有数以百计的不同步骤的过程来描述这些方法）。

1965 年，被誉为“专家系统和知识工程之父”的费根鲍姆（Feigenbaum）所领导的研究小组，开始研究专家系统，并于 1968 年研究、功第一个专家系统 DENDRAL，用于质谱仪分析有机化合物的分子结构。后来又开发出其他一些专家系统，为人工智能的应用研究做出了开创性贡献。

1969 年召开了第一届国际人工智能联合会议（International Joint Conference on AI, IJCAI），标志着人工智能作为一门独立学科登上国际学术舞台。此后，IJCAI 每两年召开一次。1970 年《人工智能国际杂志》（International Journal of AI）创刊。这些事件对开展人工智能国际学术活动和交流、促进人工智能的研究和发展起到了积极作用。

上述事件表明，人工智能经历了从诞生到成人的热烈（形成）期，已成为一门独立学科，为人工智能建立了良好的环境，打下了进一步发展的重要基础。虽然人工智能在前进的道路上仍将面临不少困难和挑战，但是有了这个基础，就能够迎接挑战，抓住机遇，推动人工智能不断发展。

3、暗淡时期（1966-1974 年）

在形成期和后面的知识应用期之间，交叠地存在一个人工智能的暗淡（低潮）期。在取得“热烈”发展的同时，人工智能也遇到一些困难和问题。

一方面，由于一些人工智能研究者被“胜利冲昏了头脑”，盲目乐观，对人工智能的未来发展和成果做出了过高的预言，而这些预言的失败，给人工智能的声誉造成重大伤害。同时，许多人工智能理论和方法未能得到通用化和推广应用，专家系统也尚未获得广泛开发。因此，看不出人工智能的重要价值。追究其因，当时的人工智能主要存在下列三个局限性：

（1）知识局限性早期开发的人工智能程序包含太少的主题知识，甚至没有知识，而且只采用简单的句法处理。例如，对于自然语言理解或机器翻译，如果缺乏足够的专业知识和常识，就无法正确处理语言，甚至会产生令人啼笑皆非的翻译。

（2）解法局限性人工智能试图解决的许多问题因其求解方法和步骤的局限性，往往使得设计的程序在实际上无法求得问题的解答，或者只能得到简单问题的解答，而这种简单问题并不需要人工智能的参与。

（3）结构局限性用于产生智能行为的人工智能系统或程序存在一些基本结构上的严重局限，如没有考虑不良结构，无法处理组合爆炸问题，因而只能用于解决比较简单的问题，影响到推广应用。

另一方面，科学技术的发展对人工智能提出新的要求甚至挑战。例如，当时认知生理学研究发现，人类大脑含有 1 个以上神经元，而人工智能系统或智能机器在现有技术条件下无法从结构上模拟大脑的功能。

此外，哲学、心理学、认知生理学和计算机科学各学术界，对人工智能的本质、理论和应用各方面，一直抱有怀疑和批评，也使人工智能四面楚歌。例如，1971 年英国剑桥大学数学家詹姆士（James）按照英国政府的旨意，发表一份关于人工智能的综合报告，声称“人工智能不是骗局，也是庸人自扰”。在这个报告影响下，英国政府削减了人工智能研究经费，解散了人工智能研究机构。在人工智能的发源地美国，连在人工智能研究方面颇有影响的 IBM，也被迫取消了该公司的所有人工智能研究。人工智能研究在世界范围内陷入困境，处于低潮，由此可见一斑。

任何事物的发展都不可能一帆风顺，冬天过后，春天就会到来。通过总结经验教训，开展更为广泛、深入和有针对性的研究，人工智能必将走出低谷，迎来新的发展时期。

4、知识应用时期（1970——1988 年）

费根鲍姆(Feigenbaum)研究小组自 1965 年开始研究专家系统，并于 1968 年研究成功第一个专家系统 DENDRAL。1972-1976 年，他们又开发成功 MYCIN 医疗专家系统，用于抗生素药物治疗。此后，许多著名的专家系统，如斯坦福国际人工智能研究中心的杜达（Duda）开发的 PROSPECTOR 地质勘探专家系统，拉特格尔大学的 CASNET 青光眼诊断治疗专家系统，MIT 的 MACSYMA 符号积分和数学专家系统，以及 RI 计算机结构设计专家系统、ELAS 钻井数据分析专家系统和 ACE 电话电缆维护专家系统等被相继开发，为工矿数据分析处理、医疗诊断、计算机设计、符号运算等提供了强有力的工具。在 1977 年举行的第五届国际人工智能联合会议上，费根鲍姆正式提出了知识工程(knowledge engineering)的概念，并预言 20 世纪 80 年代将是专家系统蓬勃发展的时代。事实果真如此，整个 80 年代，专家系统和知识工程在全世界得到迅速发展。专家系统为企业等用户赢得了巨大的经济效益。例如，第一个成功应用的商用专家系统 RI，1982 年开始在美国数字装备集团公司（DEC）运行，用于进行新计算机系统的结构设计。到 1986 年，RI 每年为该公司节省 400 万美元。到 1988 年，DEC 公司的人工智能团队开发了 40 个专家系统。更有甚者，杜珀公司已使用 100 个专家系统，正在开发 500 个专家系统。几乎每个美国大公司都拥有自己的人工智能小组，并应用专家系统或投资专家系统技术。20 世纪 80 年代，日本和西欧也争先恐后地投入对专家系统的智能计算机系统的开发，并应用于工业部门。其中，日本 1981 年发布的“第五代智能计算机计划”就是一例。在开发专家系统过程中，许多研究者获得共识，即人工智能系统是一个知识处理系统，而知识表示、知识利用和知识获取则成为人工智能系统的三个基本问题。

5、集成发展时期（1986 年至今）

到 20 世纪 80 年代后期，各个争相进行的智能计算机研究计划先后遇到严峻挑战和困难，无法实现其预期目标。这促使人工智能研究者们对已有的人工智能和专家系统思想和方法进行反思。已有的专家系统存在缺乏常识知识、应用领域狭窄、知识获取困难、推理机制单一、未能分布处理等问题。他们发现，困难反映出人工智能和知识工程的一些根本问题，如交互问题、扩展问题和体系问题等，都没有很好解决。对存在问题的探讨和对基本观点的争论，有助于人工智能摆脱困境，迎来新的发展机遇。

人工智能应用技术应当以知识处理为核心，实现软件的智能化。知识处理需要对应用领域和问题求解任务有深入的理解，扎根于主流计算环境。只有这样，才能促使人工智能研究和应用走上持续发展的道路。

20 世纪 80 年代后期以来，机器学习、计算智能、人工神经网络和行为主义等研究的深入开展，不时形成高潮。有别于符号主义的连接主义和行为主义的人工智能学派也乘势而上，获得新的发展。不同人工智能学派间的争论推动了人工智能研究和应用的进一步发展。以数理逻辑为基础的符号主义，从命题逻辑到谓词逻辑再至多值逻辑，包括模糊逻辑和粗糙集理论，已为人工智能的形成和发展做出历史性贡献，并已超出传统符号运算的范畴，表明符号主义在发展中不断寻找新的理论、方法和实现途径。传统人工智能（我们称之为 I）的数学计算体系仍不够严格和完整。除了模糊计算外，近年来，许多模仿人脑思维、自然特征和生物行为的计算方法（如神经计算、进化计算、自然计算、免疫计算和群计算等）已被引入人工智能学科。我们把这些有别于传统人工智能的智能计算理论和方法称为计算智能（computational intelligence, CI）。计算智能弥补了传统 AI 缺乏数学理论和计算的不足，更新并丰富了人工智能的理论框架，使人工智能进入一个新的发展时期。人工智能不同观点、方法和技术的集成，是人工智能发展所必需，也是人工智能发展的必然。

在这个时期,特别值得一提的是神经网络的复兴和智能真体(intelligent agent)的突起。

麦卡洛克和皮茨 1943 年提出的“似脑机器”,构造了一个表示大脑基本组成的神经元模型。由于当时神经网络的局限性,特别是硬件集成技术的局限性,使人工神经网络研究在 20 世纪 70 年代进入低潮。直到 1982 年霍普菲尔德(Hopfield)提出离散神经网络模型,1984 年又提出连续神经网络模型,促进了人工神经网络研究的复兴。布赖森(Bryson)和何(He)提出的反向传播(BP)算法及鲁梅尔哈特(Rumelhart)和麦克莱伦德(McClelland)1986 年提出的并行分布处理(PDP)理论是人工神经网络研究复兴的真正推动力,人工神经网络再次出现研究热潮。1987 年在美国召开了第一届神经网络国际会议,并发起成立了国际神经网络学会(JNNS)。这表明神经网络已置身于国际信息科技之林,成为人工智能的一个重要子学科。如果人工神经网络硬件能够在大规模集成上取得突破,那么其作用不可估量。现在,对神经网络的研究出现了 21 世纪以来的一次高潮,特别是基于神经网络的机器学习获得很大发展。近 10 年来,深度学习(deep learning)的研究逐步深入,并已在自然语言处理和人机博弈等领域获得比较广泛的应用。在深度学习的基础上,一种称为“超限学习”(extreme learning)的机器学习方法在近几年得到越来越多的应用。这些研究成果活跃了学术氛围,推动了机器学习的发展。

智能真体(以前称为智能主体)是 20 世纪 90 年代随着网络技术特别是计算机网络通信技术的发展而兴起的,并发展为人工智能又一个新的研究热点。人工智能的目标就是要建造能够表现出一定智能行为的真体,因此,真体(agent)应是人工智能的一个核心问题。人们在人工智能研究过程中逐步认识到,人类智能的本质是一种具有社会性的智能,社会问题特别是复杂问题的解决需要各方面人员共同完成。人工智能,特别是比较复杂的人工智能问题的求解也必须各个相关个体协商、协作和协调来完成的。人类社会中的基本个体“人”对应于人工智能系统中的基本组元“真体”,而社会系统所对应的人工智能“多真体系统”也就成为人工智能新的研究对象。

产业的改造与升级、智能制造和服务民生的需求,促进机器人学向智能化方向发展,一股机器人化的新热潮正在全球汹涌澎湃,席卷全世界。智能机器人已成为人工智能研究与应用的一个蓬勃发展的新领域。人工智能已获得越来越广泛的应用,深入渗透到其他学科和科学技术领域,为这些学科和领域的发展作出功不可没的贡献,并为人工智能理论和应用研究提供新的思路与借鉴。例如,对生物信息学、生物机器人学和基因组的研究就是如此。

上述这些新出现的人工智能理论、方法和技术,其中包括人工智能三大学派,即符号主义、连接主义和行为主义,已不再是单枪匹马打天下,而往往是携手合作,走综合集成、优势互补、共同发展的康庄大道。人工智能学界那种势不两立的激烈争论局面,可能一去不复返了。我们有理由相信,人工智能工作者一定能够抓住机遇,不负众望,创造更多更大的新成果,开创人工智能发展的新时期。

我国的人工智能研究起步较晚。纳入国家计划的研究(“智能模拟”)始于 1978 年;1984 年召开了智能计算机及其系统的全国学术讨论会;1986 年起把智能计算机系统、智能机器人和智能信息处理(含模式识别)等重大项目列入国家高技术研究计划;1993 年起,又把智能控制和智能化等项目列入国家科技攀登计划。进入 21 世纪后,已有更多的人工智能与智能系统研究获得各种基金计划支持,并与国家国民经济和科技发展的重大需求相结合,力求作出更大贡献。1981 年起,相继成立了中国人工智能学会(CAAI)及智能机器人专业委员会和智能控制专业委员会、全国高校人工智能研究会、中国计算机学会人工智能与模式识别专业委员会、中国自动化学会模式识别与机器智能专业委员会、中国软件行业协会人工智能协会以及智能化专业委员会等学术团体。在中国人工智能学会归属中国科学技术协会直接领导和管理之后,又有一些省市成立了地方人工智能学会,推动了我国人工智能的发展。1989 年首次召开了中国人工智能控制联合会议(CJCAI)已有约 60 部国内编著的具有知识产权的人工智能专著和教材出版,其中,本书就已出版发行 40 多万册。1982 年创刊《人工智能学报》杂志,《模式识别与人工智能》杂志和《智能系统学报》分别于 1987 年和 2006 年创刊。《智能技术学报》英文版即将创刊。2006 年 8 月,中国人工智能学会联合兄弟学会和有关部门,在北京举办了包括人工智能国际会议和中国象棋人机大战等在内的“庆祝人工智能学科诞生 50 周年”大型庆祝活动,产生了很好的影响。今年 4 月又在北京举行《全球人工智能技术大会暨人工智能 60 周年纪念活动启动仪式》,隆重而热烈地庆祝国际人工智能学科诞生 60 周年。2009 年,

中国人工智能学会牵头组织，向国家学位委员会和国家教育部提出“设置‘智能科学与技术，学位授权一级学科’”的建议，为我国人工智能和智能科学学科建设不遗余力，意义深远。2015年在中国最热门的话题和产业应该是机器人学，中国机器人学的磅礴热潮推动世界机器人产业的新一轮竞争与发展。2016年中国最为引人注目的科技应是工人智能，并出现了发展人工智能及其产业的新潮。中国的人工智能工作者，已在人工智能领域取得许多具有国际领先水平的创造性成果。其中，尤以吴文俊院士关于几何定理证明的“吴氏方法”最为突出，已在国际上产生了重大影响，并荣获2001年国家科学技术最高奖。现在，我国已有数以万计的科技人员和大学师生从事不同层次的人工智能研究与学习，人工智能研究已在我国深入开展，它必将为促进其他学科的发展和我国的现代化建设做出新的重大贡献。

1.2.3 人工智能前景方向

人工只能三大流派中的符号主义认为只要实现指名功能就可以实现人工智能；联结主义认为只要实现指心功能就可以实现人工智能；行为主义认为只要实现指物功能就可以实现人工智能。人工智能的三大流派虽然取得了很大进展，但各自也面临巨大挑战。简单地说，人工智能三大流派假设之所以能够成立的前提是指名、指物、指心功能等价。

可是这个前提成立吗？早期的人工智能研究使用经典概念，而经典概念至少具有以下5个假设：①概念的外延表示可以用经典集合表示；②概念的内涵表示存在命题表示；③指称对象的外延表示与其内涵表示名称一致；④概念表示唯一，即同一个概念的表示与个体无关，对于同一个概念，每个人表示都是一样的；⑤概念的内涵表示与外延表示在指称对象上功能等价。

可以明显看出，在上述5个假设之下，经典概念的指心、指物、指名功能是等价的，即指名意味着指物、指心。但是，日常生活中使用的概念一般并不满足经典概念的5条假设，因此也不能保证其指心、指物、指名功能等价。《周易·系辞上》中也说，“书不尽言，言不尽意”，明确指出了指名、指心与指物不一定等价。下面给出两个例子，以说明日常生活中概念的指名、指物功能并不等价。微信上曾经流传过一个著名段子：一个人说手头有一个亿，谁有项目通知一下，一起投资。不然，再晚一点，就洗手不干了。听的人以为其是指物即真的有一个亿的资金。而实际上，这个人只是在手头上写了三个字“一个亿”而已。在这儿纯粹指名，即“手头有一个亿”仅仅是符号“一个亿”而已。这个段子显然利用了概念的指名与指物不一定等价的性质。

综上所述，概念的指名、指物与指心功能在生活中并不等价，单独实现概念的一个功能并不能保证具有智能。因此，单独遵循一个学派不足以实现人工智能，现在的人工智能研究已经不再强调遵循人工智能的单一学派。很多时候会综合各个流派的技术。比如，从专家系统发展起来的知识图谱已经不完全遵循符号主义的路线了。在围棋上战胜人类顶尖棋手的AlphaGo。综合使用了三种学习算法——强化学习、蒙特卡罗树搜索、深度学习，而这三种学习算法分属于三个人工智能流派（强化学习属于行为主义，蒙特卡罗树搜索属于符号主义，深度学习属于连接主义）。无人驾驶技术同样是突破了人工智能三大流派限制的综合技术。虽然人工智能发展至今，各个流派依然在发展，也都取得了很好的进展，但是各个流派进行融合已经是大势所趋，特别是在大数据和云计算的助力下，新一代人工智能将带来社会的第四次技术革命。

然而，目前的人工智能还有很大的缺陷，其使用的知识表示还是建立在经典概念的基础之上。图灵测试的文章发表之日在1950年，当时人们对于经典概念的普适性还没有提出质疑，因此，其使用的概念是基于经典概念的。在图灵测试中，最重要的概念之一是人，包括提问者和回答者。但是什么样的人才合适是合适的提问者和回答者，是中国人还是英国人、是圣人还是智力障碍者或者装傻者，图灵测试并没有定义。如果人存在经典定义，在图灵测试里，就容易确定什么样的人作为提问者和回答者。可惜的是，人并不能用经典概念定义来定义。

经典概念的基本假设还是指心、指名与指物等价，这与人类的日常生活经验严重不符，过于简单化了。在人类的现实生活当中，概念的指名指物指心并不总是等价的。在基于经典概念的知识表示框架下，现在的机器表现有时显得极其智障，缺乏常识、缺乏理解能力，严重缺乏处理突发状况的能力。实际上，维特根斯坦在1953年出版的《哲学研究》明确提出，日常生活中使用的概念如人等是没有经典概念定义的。这实际上给图灵测试带来了很大的不确定性。因此，在经典概念表示不成立的情形下，如何进行概念表示是

一个极具挑战性的问题。

1.3 人工智能的流派

目前人工智能的主要流派有下列 3 家：

(1) 符号主义 (symbolicism)：又称为逻辑主义 (logicism)、心理学派 (psychologism) 或计算机学派 (computerism)，其原理主要为物理符号系统假设和有限合理性原理。

(2) 连接主义 (connectionism)：又称为仿生学派 (bionicsism) 或生理学派 (physiologism)，其原理主要为神经网络及神经网络间的连接机制与学习算法。

(3) 行为主义 (actionism)：又称为进化主义 (evolutionism) 或控制论学派 (cyberneticsism)，其原理为控制论及感知—动作型控制系统。

人工智能各学派对人工智能发展历史具有不同的看法。不同人工智能学派对人工智能的研究方法问题也有不同的看法。这些问题涉及：人工智能是否一定采用模拟人的智能的方法？若要模拟又该如何模拟？对结构模拟和行为模拟、感知思维和行为、对认知与学习以及逻辑思维和形象思维等问题是否应分离研究？是否有必要建立人工智能的统一理论系统？若有，又应以什么方法为基础？

1.3.1 符号主义

符号主义认为人工智能源于数理逻辑。数理逻辑从 19 世纪末起就获迅速发展；到 20 世纪 30 年代开始用于描述智能行为。计算机出现后，又在计算机上实现了逻辑演绎系统。其有代表性的成果为启发式程序 LT (逻辑理论家)，证明了 38 条数学定理，表明了可以应用计算机研究人类智能活动。正是这些符号主义者，早在 1956 年首先采用“人工智能”这个术语。后来又发展了启发式算法—>专家系统—>知识工程理论与技术，并在 20 世纪 80 年代取得很大发展。符号主义曾长期一枝独秀，为人工智能的发展作出重要贡献，尤其是专家系统的成功开发与应用，为人工智能走向工程应用和实现理论联系实际具有特别重要的意义。在人工智能的其他学派出现之后，符号主义仍然是人工智能的主流学派。这个学派的代表人物有纽厄尔、肖、西蒙和尼尔逊 (Nilsson) 等。

符号主义认为人工智能的研究方法应为功能模拟方法。通过分析人类认知系统所具备的功能和机能，然后用计算机模拟这些功能，实现人工智能。符号主义力图用数学逻辑方法来建立人工智能的统一理论体系，但遇到不少暂时无法解决的困难，并受到其他学派的否定。

1.3.2 连接主义

连接主义认为人工智能源于仿生学，特别是人脑模型的研究。它的代表性成果是 1943 年由生理学家麦卡洛克和数理逻辑学家皮茨创立的脑模型，即 MP 模型，开创了用电子装置模仿人脑结构和功能的新途径。它从神经元开始进而研究神经网络模型和脑模型，开辟了人工智能的又一发展道路。20 世纪 60-70 年代，连接主义，尤其是对以感知机 (perceptron) 为代表的脑模型的研究曾出现过热潮。由于当时的理论模型、生物原型和技术条件的限制，脑模型研究在 70 年代后期至 80 年代初期落入低潮。直到前述 Hopfield 教授在 1982 年和 1984 年发表两篇重要论文，提出用硬件模拟神经网络后，连接主义又重新抬头。1986 年鲁梅尔哈特等人提出多层网络中的反向传播 (BP) 算法。此后，连接主义势头大振，从模型到算法，从理论分析到工程实现，为神经网络计算机走向市场打下基础。现在，对 ANN 的研究热情仍然较高，但研究成果未能如预想的那样好。

连接主义主张人工智能应着重于结构模拟，即模拟人的生理神经网络结构，并认为功能、结构和智能行为是密切相关的。不同的结构表现出不同的功能和行为。已经提出多种人工神经网络结构和众多的学习算法。

1.3.3 行为主义

行为主义认为人工智能源于控制论。控制论思想早在 20 世纪 40-50 年代就成为时代思潮的重要部分，影响了早期的人工智能工作者。维纳和麦克洛 (McCloe) 等人提出的控制论和自组织系统以及钱学森等人提出的工程控制论和生物控制论，影响了许多领域。控制论把神经系统的工作原理与信息理论、控制理论、逻辑以及计算机联系起来。早期的研究重点是模拟人在控制过程中的智能行为和作用，如对自寻优、自适应、自校正、自镇定、自组织和自学习等控制论系统的研究，并进行“控制论动物”的研制。到 60-70

年代,上述这些控制论系统的研究取得一定进展,播下智能控制和智能机器人的种子,并在 80 年代诞生了智能控制和智能机器人系统。行为主义是 20 世纪末才以人工智能新学派的面孔出现的,引起许多人的兴趣。这一学派的代表作首推布鲁克斯(Brooks)的六足行走机器人,它被看作新一代的“控制论动物”,是一个基于感知—动作模式的模拟昆虫行为的控制系统。

行为主义认为人工智能的研究方法应采用行为模拟法,也认为功能、结构和智能行为是不可分的。不同行为表现出不同功能和不同控制结构。行为主义的研究方法也受到其他学派的怀疑与批判,认为行为主义最多只能创造出智能昆虫行为,而无法创造出人的智能行为。

以上三个人工智能学派将长期共存与合作,取长补短,并走向融合和集成,为人工智能的发展做出贡献。

1.4 人工智能的研究内容

人工智能学科有着十分广泛和极其丰富的研究内容。不同的人工智能研究者从不同的角度对人工智能的研究内容进行分类。例如,基于脑功能模拟、基于不同认知观、基于应用领域和应用系统、基于系统结构和支撑环境等。因此,要对人工智能研究内容进行全面和系统的介绍也是比较困难的,而且可能也是没有必要的。下面综合介绍一些得到诸多学者认同并具有普遍意义的人工智能研究的基本内容。

浩斯顿(Houston)等把认知归纳为如下 5 种类型:

- (1) 信息处理过程;
- (2) 心理上的符号运算;
- (3) 问题求解;
- (4) 思维;
- (5) 诸如知觉、记忆、思考、判断、推理、学习、想象、问题求解、概念形成和语言使用等关联活动。

人类的认知过程是非常复杂的。作为研究人类感知和思维信息处理过程的一门学科,认知科学(或称思维科学)就是要说明人类在认知过程中是如何进行信息加工的。认知科学是人工智能的重要理论基础,涉及非常广泛的研究课题。除了浩斯顿提出的知觉、记忆、思考、学习、语言、想象、创造、注意和问题求解等关联活动外,还会受到环境、社会和文化背景等方面的影响。人工智能不仅要研究逻辑思维,而且还要深入研究形象思维和灵感思维,使人工智能具有更坚实的理论基础,为智能系统的开发提供新思想和新途径。

1、知识表示

知识表示、知识推理和知识应用是传统人工智能的三大核心研究内容。其中,知识表示是基础,知识推理实现问题求解,而知识应用是目的。知识表示是把人类知识概念化、形式化或模型化。一般地,就是运用符号知识、算法和状态图等来描述待解决的问题。已提出的知识表示方法主要包括符号表示法和神经网络表示法两种。

2、专家系统

人工智能能否获得广泛应用是衡量其生命力和检验其生存力的重要标志。20 世纪 70 年代,正是专家系统的广泛应用,使人工智能走出低谷,获得快速发展。后来的机器学习和近年来的自然语言理解应用研究取得重大进展,又促进了人工智能的进一步发展。当然,应用领域的发展是离不开知识表示和知识推理等基础理论以及基本技术的进步的。

3、机器感知

机器感知就是使机器具有类似于人的感觉,包括视觉、听觉、力觉、触觉、嗅觉、痛觉、接近感和速度感等。其中,最重要的和应用最广的要算机器视觉(计算机视觉)和机器听觉。机器视觉要能够识别与理解文字、图像、场景以至人的身份等;机器听觉要能够识别与理解声音和语言等。机器感知是机器获取外部信息的基本途径。要使机器具有感知能力,就要为它安上各种传感器。机器视觉和机器听觉已催生了人工智能的两个研究领域——模式识别和自然语言理解或自然语言处理。实际上,随着这两个研究领域的进展,它们已逐步发展成为相对独立的学科。

4、机器思维

机器思维是对传感信息和机器内部的工作信息进行有目的的处理。要使机器实现思维，需要综合应用知识表示、知识推理、认知建模和机器感知等方面的研究成果，开展如下各方面的研究工作：

- (1) 知识表示，特别是各种不确定性知识和不完全知识的表示。
- (2) 知识组织、积累和管理技术。
- (3) 知识推理，特别是各种不确定性推理、归纳推理、非经典推理等。
- (4) 各种启发式搜索和控制策略。
- (5) 人脑结构和神经网络的工作机制。

5、机器学习

机器学习是继专家系统之后人工智能应用的又一重要研究领域，也是人工智能和神经计算的核心研究课题之一。现有的计算机系统和人工智能系统大多数没有什么学习能力，至多也只有非常有限的学习能力，因而不能满足科技和生产提出的新要求。学习是人类具有的一种重要智能行为。机器学习就是使机器（计算机）具有学习新知识和新技术，并在实践中不断改进和完善的能力。机器学习能够使机器自动获取知识，向书本等文献资料和与人交谈或观察环境进行学习。

6、机器行为

机器行为系指智能系统（计算机，机器人）具有的表达能力和行动能力，如对话、描写、刻画以及移动、行走、操作和抓取物体等。研究机器的拟人行为是人工智能的高难度的任务。机器行为与机器思维密切相关，机器思维是机器行为的基础。

7、计算机智能系统

上述直接的实现智能研究，离不开智能计算机系统或智能系统，离不开对新理论、新技术和新方法以及系统的硬件和软件支持。需要开展对模型、系统构造与分析技术、系统开发环境和构造工具以及人工智能程序设计语言的研究。一些能够简化演绎、机器人操作和认知模型的专用程序设计以及计算机的分布式系统、并行处理系统、多机协作系统和各种计算机网络等的发展，将直接有益于人工智能的开发。

1.5 人工智能的典型应用领域

在大多数学科中存在着几个不同的研究领域，每个领域都有其特有的感兴趣的研究课题、研究技术和术语。在人工智能中，这样的领域包括自然语言处理、自动定理证明、自动程序设计、智能检索、智能调度、机器学习、机器人学、专家系统、智能控制、模式识别、视觉系统、神经网络、agent、计算智能、问题求解、人工生命、人工智能方法和程序设计语言等。在过去 60 年中，已经建立了一些具有人工智能的计算机系统；例如，能够求解微分方程的，下棋的，设计分析集成电路的，合成人类自然语言的，检索情报的，诊断疾病以及控制太空飞行器、地面移动机器人和水下机器人的具有不同程度人工智能的计算机系统。

值得指出的是，正如不同的人工智能子领域不是完全独立的一样，这里简介的各种智能特性也不是互不相关的。把它们分开来介绍只是为了便于指出现有的人工智能程序能够做些什么和还不能做什么。大多数人工智能研究课题都涉及许多智能领域。

1、搜索和问题消解

人工智能的第一个大成就是发展了能够求解难题的下棋（如国际象棋）程序。在下棋程序中应用的某些技术，如向前看几步，并把困难的问题分成一些比较容易的子问题，发展成为搜索和问题消解（归约）这样的人工智能基本技术。今天的计算机程序能够下锦标赛水平的各种方盘棋、十五子棋、中国象棋和国际象棋，并取得前面提到的计算机棋手战胜国际和国家象棋冠军的成果。另一种问题求解程序把各种数学公式符号汇编在一起，其性能达到很高的水平，并正在为许多科学家和工程师所应用。有些程序甚至还能够用经验来改善其性能。

如前所述，这个问题中未解决的问题包括人类棋手具有的但尚不能明确表达的能力，如国际象棋大师们洞察棋局的能力。另一个未解决的问题涉及问题的原概念，在人工智能中叫做问题表示的选择。人们常常能够找到某种思考问题的方法从而使求解变易而解决该问题。到目前为止，人工智能程序已经知道如何考虑它们要解决的问题，即搜索解答空间，寻找较优的解答。