

人工智能基础绪论

回溯人类社会的发展历程，工业革命、信息革命以及当下如火如荼的人工智能革命，依次推动着时代的车轮滚滚向前，深刻地改变了人类的生产与生活方式。

工业革命率先拉开帷幕，其核心在于体力的解放。18 世纪 60 年代至 19 世纪中期，第一次工业革命以蒸汽机的广泛应用为标志，纺织机、蒸汽机车等机械设备相继问世，开启了机械化生产的先河。随后的第二次工业革命，电力成为主导力量，电动机、内燃机等发明推动了工业化进程，生产方式更加多样化。到了第三次工业革命，计算机技术的融入实现了生产的自动化与精准化，现代工业体系逐渐成型。

紧随其后的是信息革命，它似乎预示着脑力解放的曙光。20 世纪中叶以来，计算机的诞生与普及极大地提高了信息处理能力。1946 年，世界上第一台计算机“ENIAC”问世，开启了计算机时代。此后，互联网的兴起更是让信息传播速度呈指数级增长，人们获取知识的途径变得前所未有的便捷。然而，信息化的浪潮也带来了过度信息化的隐忧，信息过载让人们在海量数据中难以迅速筛选出真正有价值的内容。

如今，人工智能革命正以前所未有的速度和规模席卷全球，它不仅是脑力劳动自动化的体现，更是对人类智能的延伸与拓展。从 20 世纪 50 年代达特茅斯会议确立人工智能学科至今，人工智能已经取得了长足的发展，2016 年 AlphaGo 的出现标志着人工智能技术的一个重要转折点。如今，人工智能在医疗、金融、交通等多个领域发挥着重要作用，推动着产业和社会向智能化、自动化的方向快速发展。

1. 什么是人工智能

1.1 什么是智能

1.1.1 智能特征

智能的本质是一个复杂而多维的概念，涵盖了多个关键能力，这些能力共同构成了智能的**核心特征**。

首先，**模仿**是智能的重要表现之一，通过模仿，智能体能够学习和复制他人的行为或模式，从而快速适应新环境或任务。例如，儿童通过模仿成人的语言和行为来学习社会规范和交流方式。

其次，**理解**是智能的另一个核心要素。理解不仅仅是对信息的表面认知，而是能够深入把握信息的内在含义和逻辑关系。例如，理解一段文字不仅需要知道每个单词的意思，还需要领会其背后的意图和情感。这种深层次的理解能力是智能体进行有效决策和交流的基础。

学习是智能发展的关键驱动力。通过学习，智能体能够不断积累知识和经验，从而提升自身的能力和适应性。学习可以是显式的，如通过教育和培训，也可以是隐式的，如通过实践和经验的积累。例如，一个人通过学习数学公式来解决复杂的计算问题，或者通过实践来掌握一门新的技能。

思维是智能的内在表现形式，包括逻辑推理、问题解决和创造性思考等。思维能力使智能体能够对信息进行加工和处理，从而形成新的知识和见解。例如，科学家通过逻辑推理和实验验证来发现新的科学规律，艺术家通过创造性思考来创作独特的艺术作品。

推理是智能的重要工具，使智能体能够从已知信息中推导出未知结论。推理可以是演绎推理，即从一般到具体的推理，也可以是归纳推理，即从具体到一般的推理。例如，医生通过患者的症状和病史进行推理，从而诊断出疾病。

最后，**行为**是智能的外在表现，反映了智能体在特定环境中的适应能力和决策能力。行为不仅仅是对外界刺激的简单反应，而是基于对环境的理解和对目标的追求。例如，动物通过行为来寻找食物、躲避天敌和繁衍后代，人类通过行为来实现个人目标和社会价值。

综上所述，智能的本质在于其多维度的能力，包括模仿、理解、学习、思维、推理和行为，这些能力相互交织，共同构成了智能体在复杂环境中生存和发展的基础。

1.1.2 关于智能的理论

在探讨智能这一概念时，可以从多个理论角度来深入理解其含义。首先，思维理论认为智能的核心在于思维，它强调通过研究思维的规律和方法，能够揭开智能背后的本质。这种理论视角将智能视为一种高级的认知活动，它涉及逻辑推理、问题解决以及决策制定等过程。思维理论试图解释智能是如何在大脑中运作的，以及它是如何影响我们对世界的理解和行动的。

其次，知识阈值理论为我们提供了另一种理解智能的途径。这一理论认为，智能行为的展现与个体所拥有的知识量以及知识的一般化程度密切相关。换句话说，智能不仅仅依赖于知识的积累，更依赖于如何将这些知识进行抽象化和泛化，以便在面对复杂问题时能够迅速地找到解决方案。这种理论强调了知识的组织和应用能力，认为智能个体能够在庞大的知识空间中有效地导航，并且能够灵活地运用知识来应对各种挑战。

最后，进化理论为我们理解智能提供了生物学和环境适应性的视角。由麻省理工学院（MIT）的布鲁克（R.A.Brook）教授所提出的这一理论，特别强调了智能生物对外界事物的感知能力以及对动态变化环境的适应能力。这种理论视角认为，智能是生物进化过程中为了更好地生存和繁衍而发展出来的能力。它不仅仅局限于人类，也适用于其他动物和可能存在的智能生命形式。进化理论认为，智能的形成和发展是与生物体所处的环境紧密相连的，智能生物必须能够感知环境变化，并作出相应的适应性反应。

这三种理论观点，从思维、知识和进化三个不同的维度，为我们全面地阐释了智能的内涵。它们为我们提供了多维度的视角，帮助我们更深入地理解智能这一复杂现象。通过综合这些理论，我们可以更全面地认识到智能不仅仅是解决问题的能力，它还涉及到认知、学习、适应以及与环境的互动等多个方面。这些理论为我们研究智能提供了丰富的理论基础，也为智能技术的发展和应用的指明了方向。

1.1.3 什么是智能

智能（Intelligence）是一个复杂而多维的概念，涵盖了知识与智力的总和。具体来说：

智能是知识与智力的总和：智能不仅仅是知识的积累，还包括运用知识的能力。知识是智能行为的基础，而智力则是获取和运用知识的能力，它来自于人脑的思维活动。

知识是智能行为的基础：知识是智能行为的基础，没有知识的积累，智能行为就无法有效展开。知识的广度和深度直接影响智能的表现。

智力是获取知识、运用知识的能力：智力是人脑通过思维活动获取和运用知识的能力，包括学习、理解、推理、解决问题等。智力的高低决定了一个人在面对新问题时的适应能力和解决能力。

对外界事物的感知能力是智能的重要基础及组成部分：感知能力使我们能够感知外界事物，获取信息，从而为智能行为提供基础。感知能力的强弱直接影响知识的获取和智力的发挥。

1.1.4 人类大脑是如何实现智能的

人类大脑是实现智能的核心器官，但其奥秘至今仍有许多未解之谜。目前，人类对大脑的理解仍然非常有限：

两大难题之一：宇宙起源、人脑奥秘。宇宙起源和人脑奥秘被认为是科学界的两大难题。人脑的复杂性和精妙性远超我们的想象，其神经网络的结构和功能至今仍有许多未解之谜。

目前对人脑奥秘知之甚少：尽管现代科学技术已经取得了显著进步，但对人脑的了解仍然非常有限。人脑的神经元数量庞大，神经网络的连接复杂，这些都使得人脑的研究充满了挑战。

1.2 什么是人工智能

1.2.1 人工智能的定义

关于“人工智能”的含义，早在其被正式提出之前，英国数学家图灵（A. M. Turing）就有所提及。1950 年，图灵发表了《计算机与智能》（Computing Machinery and Intelligence）这一论文，论述并提出了著名的“图灵测试”，形象地阐述了人工智能的概念以及机器应达到的智能标准。在该论文里，图灵指出不要纠结于机器是否能思维，而是看它能否通过这样的测试：让人和机器分别处于两个房间，二者可通话，但互相看不到。若通过对话，人无法分辨对方是人还是机器，那就可认为该机器达到了人类智能水平。不过，很多人觉得图灵测试只反映结果，未涉及思维过程，即便机器通过测试，也不能认定其有智能。

那么，人工智能到底是什么呢？美国斯坦福大学的尼尔逊教授将其定义为：“人工智能是关于知识的学科——怎样表示知识以及怎样获取知识并使用知识的科学。”美国麻省理工学院的温斯顿教授则认为：“人工智能就是研究如何让计算机做过去只有人才能做的智能工作。”这些定义反映了人工智能学科的基本思想和内容，即人工智能研究人类智能活动的规律，构建具有一定智能的人工系统，探究如何让计算机完成以往需人类智力才能完成的工作，也就是研究如何运用计算机软硬件模拟人类某些智能行为的基本理论、方法和技术。

然而，关于人工智能的形式化定义，目前尚无法无统一。这是因为人工智能的严格定义取决于对智能的定义，而智能本身也没有严格定义。这使得人工智能的定义成为开放性问题，其定义会随着研究深入和技术发展而不断演变。

一般解释为，人工智能是用人工方法在机器（计算机）上实现的智能，即机器智能。简单来说，人工智能借助计算机程序和算法，使机器模拟人类智能行为，如学习、推理、解决问题等。

从学科角度看，人工智能是一门研究如何构建智能机器或智能系统，使其能模拟、延伸、扩展人类智能的学科。它涉及计算机科学、心理学、哲学、神经科学等多领域，是跨学科的研究领域。

从智能能力角度而言，人工智能是智能机器所执行的通常与人类智能有关的智能行为，如判断、推理、证明、识别、感知、理解、通信、设计、思考、规划和问题求解等思维活动。这些能力让人工智能能在各种复杂环境中展现类似人类的智能行为。

人工智能的应用极为广泛，涵盖日常生活到工业生产的各个方面。例如，智能家居系统能通过学习用户习惯自动调节温度和照明；医疗领域中，人工智能可辅助医生进行疾病诊断和治疗方案制定；交通领域里，自动驾驶技术正逐步实现，让出行更安全高效。

1.2.2 人工智能的分类

根据人工智能是否能够真正实现推理、思考和解决问题的能力，人工智能可以被划分为两个主要类别：弱人工智能和强人工智能。

1. 弱人工智能

定义：弱人工智能，也被称为窄人工智能，是指那些设计用来执行特定任务的智能机器，它们并不具备真正的推理和解决问题的能力。

特点：这些机器虽然在执行特定任务时表现得像是具有智能，但实际上它们并不具备真正的智能，也没有自主意识。它们通常只能在预设的参数和规则下工作，缺乏对任务背后深层次理解和思维的能力。

例子：一个典型的例子是智能家居系统，它可以根据一系列预设的规则自动调节家中的温度和照明。尽管它能够执行这些任务，但智能家居系统并不真正理解这些行为的意义，它只是按照编程好的程序来执行命令。

2. 强人工智能

定义：强人工智能，也被称作通用人工智能，指的是那些能够真正进行推理、思考和解决问题的智能机器。

特点：这类机器被认为具有知觉和自我意识，能够进行复杂的认知任务，包括理解、学习和适应。强人工智能可以进一步细分为两类：

- **类人：**这类机器的思考和推理方式类似于人类，它们能够模拟人类的思维过程。
- **非类人：**这类机器发展出了与人类截然不同的知觉和意识，它们使用不同于人类的推理方式来处理信息和解决问题。

例子：目前，强人工智能更多地存在于理论研究和科幻作品中。然而，一些先进的 AI 系统正在逐步向这个目标迈进。例如，一些高级的聊天机器人已经能够进行复杂的对话，并在一定程度上识别和模拟人类的情感。

从强人工智能和弱人工智能的视角出发，当下人工智能的发展形势具备如下特点：

在强人工智能方面：

- (1) 技术突破与展望：强人工智能（AGI）旨在达到与人类等同甚至超越人类的智能水准，拥有强大的逻辑推理、复杂思维以及自主学习能力。在未来 30 年里，量子计算、神经网络、自然语言处理、计算机视觉等领域的突破，将为强人工智能的达成给予有力的技术支持。预计到 2050 年，强人工智能技术能够在诸多领域广泛应用，进而成为社会发展的关键驱动力。
- (2) 社会影响与挑战：强人工智能会深度改变人类社会，使生产方式、生活方式以及社会结构等发生变革。它或许会重构社会结构，虽然部分岗位可能被替代，但也会催生新型就业机会与商业模式。并且，强人工智能可能拥有类人情感和意识，这会改变人与机器的关系，从而引发诸如伦理道德、社会公正等重要问题。

在弱人工智能方面：

- (1) 应用拓展与深化：弱人工智能聚焦于特定任务或功能，近年来在众多领域发挥了巨大作用。其借助智能算法与大数据分析来提升工作效率、削减成本并提高服务质量。未来，弱人工智能将在金融、医疗、制造业、物流等更多领域得到广泛运用，推动各行各业数字化转型。
- (2) 人机协作与智能化产品普及：随着弱人工智能的发展，人机协作会更为紧密，人们将更多地与智能系统交互来共同完成任务。例如在医疗领域，医生可依靠弱人工智能系统辅助诊断和制定治疗方案。同时，智能家居产品、智能手机、智能汽车等智能化产品将更加普及，给人们的生活增添更多便利。
- (3) 数据安全与隐私保护：弱人工智能的发展也引出了数据安全与隐私保护这一重要话题。随着人工智能应用的普及，数据隐私和安全问题愈发严峻，有必要加强相关法律法规的制定与执行，以保护用户的隐私和数据安全。

总体而言，当前人工智能的发展正处于从弱人工智能迈向强人工智能的过渡阶段，弱人工智能在各个领域广泛应用且不断深入，强人工智能则在技术研发和社会影响方面呈现出巨大的潜力与挑战。

1.3 人工智能的发展历史

人工智能（Artificial Intelligence, AI）自诞生以来，已经经历了 69 年的发展历程。回顾历史，可以根据不同时期的主要特征，将人工智能的产生与发展过程划分为几个不同的阶段。

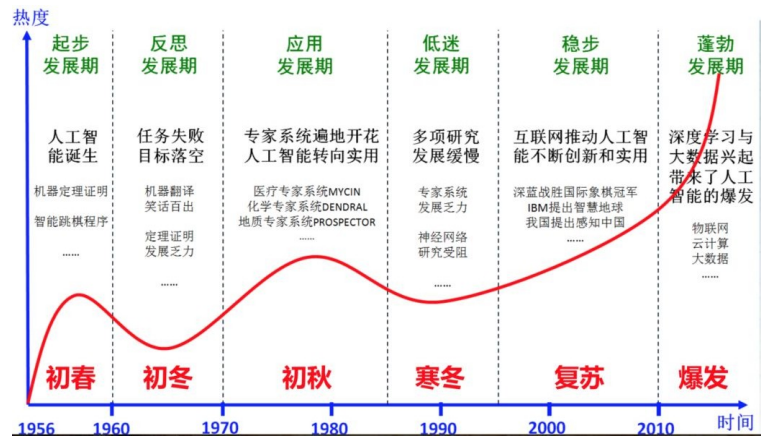


图 1 人工智能发展的不同阶段

1. 孕育期（1956 年以前）

在这一阶段，人工智能的概念尚未正式提出，但相关的思想和技术已经开始萌芽。科学家们开始思考机器是否能够模拟人类的智能行为，这为后来人工智能的诞生奠定了基础。

2. 形成期（1956 - 1970 年）

1956 年，达特茅斯会议（Dartmouth Conference）标志着人工智能作为一门独立学科の正式诞生。在这一时期，科学家们提出了许多基本概念和理论，如图灵测试、神经网络等，并开始尝试开发一些简单的智能程序。

3. 暗淡期（1966 - 1974 年）

这一阶段，人工智能的发展遇到了一些困难和挫折。由于技术限制和计算能力的不足，许多早期的乐观预期未能实现，导致研究资金和兴趣的减少，人工智能进入了所谓的“寒冬”时期。

4. 知识应用期（1970 - 1988 年）

随着计算机技术的进步，人工智能开始在特定领域取得一些突破，如专家系统的开发。这一时期，人工智能的研究重点转向了知识表示和推理，试图通过模拟人类专家的决策过程来解决实际问题。

5. 集成发展期（1986 - 2006 年）

在这一阶段，人工智能技术逐渐成熟，开始向多个方向发展。机器学习、数据挖掘、自然语言处理等领域的研究取得了显著进展，人工智能的应用范围也不断扩大，涵盖了医疗、金融、交通等多个行业。

6. 深度学习异军突起（2006 年以来）

2006 年以后，深度学习技术的出现标志着人工智能进入了一个新的发展阶段。深度学习通过构建多层神经网络，能够自动从大量数据中学习特征，极大地提高了人工智能的性能。这一技术在图像识别、语音识别、自然语言处理等领域取得了突破性成果，推动了人工智能的广泛应用和快速发展。

7. 大模型降临时代(2017 年以来)

2017 年，Google 推出了基于自注意力机制的 Transformer 架构，为大规模模型预训练算法奠定了基础，开启了 AI 新纪元。2022 年 11 月，ChatGPT 问世，凭借自然语言交互和内容生成能力在全球引起轰动。2024 年 12 月，DeepSeek 崛起，推动人工智能技术进入“普惠”时代，让更多人受益。

1.3.1 孕育期(1956 年以前)

在人工智能问世之前，一系列关键的科技成果已经为它的诞生奠定了坚实的思想、理论基础和技术条件，相关学科也取得了显著的研究成果：

- 数理逻辑：数理逻辑的发展为人工智能的逻辑推理和形式化方法提供了基础。
- 控制论：控制论研究系统的控制和通信，为人工智能的控制系统设计提供了理论支持。
- 计算理论：计算理论的发展为人工智能的算法设计和计算能力提供了基础。
- 神经网络模型：神经网络模型的研究为人工智能的机器学习和深度学习提供了理论基础。
- 电子数字计算机：电子数字计算机的发明为人工智能的实现提供了硬件支持。

在我国古代，先人们对智能机器的遐想和创造充满了丰富的想象力和创造力。他们不仅在文学作品中描绘了各种神奇的机械装置，而且在实际生活中也尝试着制造出一些能够辅助人类工作的机械装置。

- 三千多年前的古代机器人：据《列子·汤问》记载，传说周穆王在西巡途中，遇到一位名叫偃师的能工巧匠。偃师献上一个会歌舞表演的“假倡”，“揜（抑）其颐则歌合律，捧其手则舞应节，千变万化，惟意所适”。
- 两千多年前的侦察机：据《墨子·鲁问》记载，“公输子（鲁班）削竹木以为鹊”，“三日不下”。他还造了能载人的大木鸢，在战争中担任侦察的任务。
- 指南车：东汉张衡发明的指南车，能够指示方向，是古代机械工程的杰出代表。
- 木牛流马：据传由诸葛亮发明，木牛流马是一种能够自动行走的木制机械，用于运输物资。

在西方世界，众多卓越的科学家涌现而出，他们凭借持续的努力和创新思维，为人类社会的发展做出了巨大贡献。这些科学家不仅在其专业领域实现了突破性的成就，而且他们的理论和发现也对人工智能领域产生了深远的影响。

- 亚里士多德（Aristotle，公元前 384 - 322）：古希腊伟大的哲学家和思想家，创立了演绎法。他提出的三段论至今仍然是演绎推理的最基本出发点。
- 弗兰西斯·培根（F. Bacon，1561 - 1626）：英国哲学家、作家和科学家，系统地提出了归纳法，成为和亚里士多德的演绎法相辅相成的思维法则。
- 莱布尼茨（G.W. Leibnitz，1646 - 1716）：德国数学家和哲学家，把形式逻辑符号化，奠定了数理逻辑的基础，从而能够对人的思维进行运算和推理。
- 布尔（Boole，1815 - 1864）：英国数学家、逻辑学家，实现了莱布尼茨的思维符号化和数学化的思想，提出了一种崭新的代数系统——布尔代数。

以下这些科学家为人工智能的产生做出了更直接的贡献，他们的工作和研究不仅推动了人工智能领域的发展，而且对整个科技界产生了深远的影响。

图灵(A.M. Turing, 1912 - 1954)：英国数学家，1936 年创立了自动机理论，提出了图灵机的概念。图灵机是一个理论计算机模型，为计算机科学的发展奠定了基础。图灵的工作不仅推动了计算机科学的发展，也为后来的人工智能研究提供了重要的理论支持。

莫克利(J.W. Mauchly, 1907 - 1980)：美国数学家，电子数字计算机的先驱。他与研究生埃克特 (J.P. Eckert) 合作，于 1946 年研制成功了世界上第一台通用电子计算机 ENIAC。ENIAC 的出现标志着电子计算机时代的到来，为人工智能的发展提供了强大的计算能力。

麦克洛奇(W. McCulloch) 和皮兹 (W. Pitts)：美国神经生理学家，1943 年建成了第一个神经网络模型（MP 模型）。他们提出，适当的网络能够学习，这一理论为后来的人工神经网络研究奠定了基础。麦克洛奇和皮兹的工作揭示了神经网络的基本原理，为人工智能的机器学习领域提供了重要的理论支持。

香农(C.E. Shannon, 1916 - 2001)：美国数学家，1948 年发表了《通讯的数学理论》，标志着信息论的诞生。香农的工作为信息的量化、传输和处理提供了理论基础，对人工智能中的信息处理和数据传输具有重要意义。

维纳(N. Wiener, 1874 - 1956)：美国著名数学家，控制论的创始人。1948 年创立了控制论，研究系统的信息变换和控制过程。维纳的工作为人工智能的行为主义学派的形成奠定了基础，对智能系统的控制和反馈机制研究具有重要影响。

1.3.2 人工智能的诞生

1956 年夏天，美国达特茅斯（Dartmouth）大学举办了一场为期两个月的历史性聚会，这次聚会标志着人工智能（AI）的正式诞生。在这次聚会上，麦卡锡首次正式使用了“AI”这一术语。

这次聚会的发起人包括四位杰出的科学家：

- 麦卡锡 (J. McCarthy): 数学家和计算机专家，他在人工智能领域做出了开创性的贡献。
- 明斯基 (M. L. Minsky): 哈佛大学数学家和神经学家，他在人工智能和认知科学领域有着深远的影响。
- 罗切斯特 (N. Rochester): IBM 公司信息中心负责人，他在计算机科学和人工智能领域有着丰富的经验。
- 香农 (C. E. Shannon): 贝尔实验室信息部数学家和信息学家，他被誉为“信息论之父”，对人工智能的发展起到了重要的推动作用。

此外，这次聚会还邀请了多位来自不同机构的著名科学家：

- 莫尔 (T. More) 和 塞缪尔 (A. L. Samuel): 来自 IBM 公司，他们在人工智能和机器学习领域有着重要的贡献。
- 塞尔弗里奇 (O. Selfridge) 和 索罗蒙诺夫 (R. Solomonoff): 来自麻省理工学院（MIT），他们在模式识别和机器学习方面做出了开创性的工作。
- 纽厄尔 (A. Newell): 来自兰德（RAND）公司，美国，他在人工智能和认知科学领域有着深远的影响。
- 西蒙 (H. A. Simon): 来自卡内基梅隆（CMU）大学，他在人工智能和决策理论方面做出了重要贡献。

这次聚会不仅汇聚了当时最杰出的科学家，还为人工智能的发展奠定了坚实的基础。与会者们共同探讨了人工智能的未来发展方向，提出了许多重要的概念和研究课题，为后续的人工智能研究提供了宝贵的指导。

1956 Dartmouth Conference: The Founding Fathers of AI

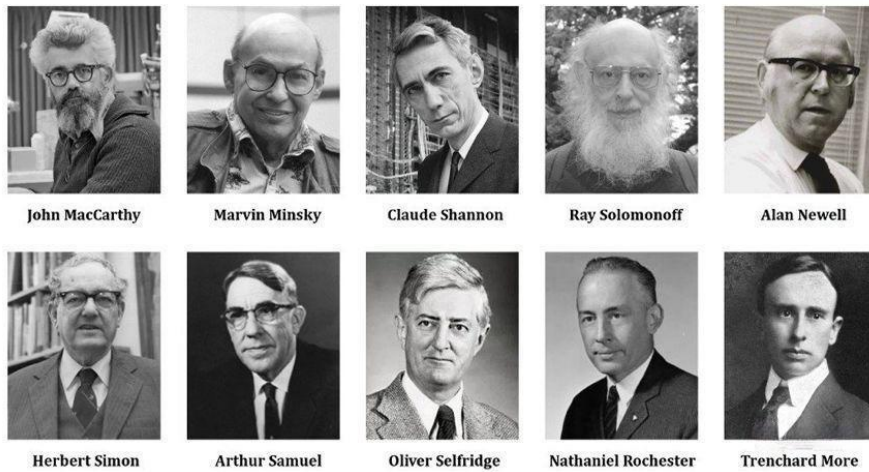


图 2 AI 的创始人们

1.3.3 形成期(1956-1970 年)

在这一时期，人工智能无论是在理论层面还是应用领域都取得了显著的成就，为后续的发展奠定了坚实的基础，这一时期也被称为人工智能发展的第一次高潮。

- **心理学小组：**1957 年，西蒙和纽厄尔等人的心理学小组研制了一个称为逻辑理论机（Logic Theory Machine，简称 LT）的数学定理证明程序。这一程序能够自动证明数学定理，展示了机器在逻辑推理方面的潜力。1960 年，该小组又研制了通用问题求解（General Problem Solving，简称 GPS）程序。GPS 的设计模仿了人类解决问题的过程，不依赖于具体领域知识，能够处理有限类别的问题。它通过程序决定子目标及可能采取的行动的次序，与人类解决问题的方式类似。因此，GPS 是第一个实现了“像人一样思考”方法的程序。
- **IBM 工程小组：**1956 年，塞缪尔在 IBM 704 计算机上研制成功了具有自学习、自组织和自适应能力的西洋跳棋程序。这一程序能够从棋谱中学习，也可以在下棋过程中积累经验、提高棋艺。通过不断学习，该程序在 1959 年击败了塞缪尔本人。1962 年，该程序又击败了一个州的冠军，展示了机器在游戏领域的智能潜力。
- **MIT 小组：**1960 年，麦卡锡研制了人工智能语言 LISP。LISP 是一种专门用于人工智能研究的编程语言，具有强大的符号处理能力，为后续的人工智能研究提供了重要的工具。

其他开创性贡献：

- 1958 年，美籍华人数理逻辑学家王浩在 IBM - 704 计算机上仅用了 3 - 5 分钟就证明了《数学原理》命题演算全部 220 条定理，展示了机器在数学领域的强大能力。
- 1965 年，鲁宾逊（J.A. Robinson）提出了归结（消解）原理，这一原理在逻辑推理和自动定理证明中具有重要应用。
- 1965 年，费根鲍姆（E. A. Feigenbaum）开始研究化学专家系统 DENDRAL，用于质谱仪分析有机化合物的分子结构。DENDRAL 是早期专家系统的重要代表，展示了人工智能在科学领域的应用潜力。

重要事件

- 1969 年，第一届国际人工智能联合会议（International Joint Conference on AI, IJCAI）召开，标志着人工智能作为一门独立学科登上了国际学术舞台。此后，IJCAI 每两年召开一次，成为人工智能领域的重要学术会议。
- 1970 年，《International Journal of AI》创刊，为人工智能研究提供了重要的学术交流平台，促进了该领域的快速发展。

1.3.4 暗淡期(1966-1974 年)

在人工智能的早期发展中，尽管取得了一些显著的成果，但也遭遇了诸多挫折和困境，进入了所谓的“暗淡期”。然而，这些挫折也为后续的研究提供了宝贵的经验和教训，为人工智能的进一步发展奠定了基础。

1. 失败的预言给人工智能的声誉造成重大伤害

60 年代初的西蒙预言：当时，西蒙等人曾预言，10 年内计算机将成为世界冠军，将证明一个未发现的数学定理，将能谱写出具有优秀作曲家水平的乐曲，大多数心理学理论将在计算机上形成。然而，这些预言并未实现，给人工智能的声誉带来了负面影响。

2. 挫折和困境

- 在博弈方面：塞缪尔的下棋程序在与世界冠军对弈时，5 局败了 4 局，未能达到预期的效果。
- 在定理证明方面：鲁宾逊归结法的能力有限，当用归结原理证明两个连续函数之和还是连续函数时，推了 10 万步也没证出结果，显示出在复杂问题上的局限性。
- 在问题求解方面：对于不良结构，会产生组合爆炸问题，导致计算复杂度急剧增加，难以找到有效的解决方案。

- 在机器翻译方面：发现并不那么简单，甚至会闹出笑话。例如，把“心有余而力不足”的英语句子翻译成俄语，再翻译回来时竟变成了“酒是好的，肉变质了”，显示出机器翻译在语义理解上的不足。
- 在神经生理学方面：研究发现人脑有 10^{11} - 10^{12} 以上的神经元，在现有技术条件下用机器从结构上模拟人脑是根本不可能的，这使得神经网络的研究遇到了瓶颈。
- 在其他方面：人工智能也遇到了不少问题。在英国，剑桥大学的詹姆斯教授指责“人工智能研究不是骗局，也是庸人自扰”，这使得人工智能研究的形势急转直下，陷入困境，落入低谷。

3. Minsky 的批评

1969 年 M. Minsky 和 S. Papert 在《感知机》一书中的批评：他们指出了感知机无法解决异或（XOR）问题的缺陷，并表示出对这方面研究的悲观态度，使得神经网络的研究从兴起期进入了停滞期。

该批评对人工智能的发展造成了重要的影响，在以后的二十年，感知机的研究方向被忽视，神经网络的研究进展缓慢。基于符号的知识表示成为主流，人工智能研究更多地集中在符号处理和逻辑推理上。基于逻辑的推理成为主要研究方向，人工智能的研究重点转向了知识表示和推理技术。

4. 当时的人工智能存在三个方面的局限性

- **知识局限性：**早期开发的人工智能程序中包含了太少的主题知识，甚至没有知识，而且只采用简单的句法处理，缺乏对语义和背景知识的理解。
- **算法局限性：**求解方法和步骤的局限性使得设计的人工智能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，显示出算法在处理复杂问题时的不足。
- **结构局限性：**用于产生智能行为的人工智能系统或程序在一些基本结构上严重局限，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，影响到人工智能系统的推广应用。

1.3.5 知识应用期(1970-1988 年)

在 1970 到 1988 年这个阶段，人工智能从理论研究转向了实际应用，特别是以知识为中心的研究取得了重大突破。这个时期的核心是专家系统，它通过模拟专家的知识 and 经验来解决实际问题，标志着 AI 从一般性的思维规律探讨转向了专门知识的应用。

一个典型的例子是 1972 年费根鲍姆开始研究的 MYCIN 医疗专家系统，1976 年成功研制。MYCIN 能够协助内科医生诊断细菌感染疾病，并提供最佳治疗处方。它不仅提高了诊断的准确性，还减轻了医生的工作负担。技术上，MYCIN 解决了知识表示、不精确推理、搜索策略、人机交互、知识获取和专家系统基本结构等关键技术问题，为后来的专家系统发展奠定了基础。



图 3 费根鲍姆(E. A. Feigenbaum)

与此同时，计算机视觉、机器人技术、自然语言理解和机器翻译等领域也取得了显著进展，丰富了人工智能的应用场景，推动了技术的创新。

在开发专家系统的过程中，研究者们逐渐认识到，人工智能系统本质上是一个知识处理系统。其核心问题包括：

知识表示：如何将人类的知识转化为计算机可以处理的形式。

知识利用：如何在系统中有效使用这些知识进行推理和决策。

知识获取：如何从专家或其他来源获取这些知识，以支持系统的持续更新和扩展。

尽管专家系统取得了显著的进展，但在实际应用中暴露出了一些问题：

应用领域狭窄：专家系统通常只能在特定领域发挥作用，缺乏通用性。

缺乏常识性知识：系统在处理需要常识判断的问题时表现不佳，限制了其适应性和灵活性。

知识获取困难：从专家或其他来源获取知识的过程复杂且耗时，影响了系统的扩展和更新。

推理方法单一：系统在处理复杂问题时，单一的推理方法显得力不从心。

没有分布式功能：缺乏分布式处理能力，限制了系统的扩展性和实用性。

不能访问现存数据库：无法直接利用现有的数据资源，影响了系统的效率和准确性。

1.3.6 集成发展期(1986-2006 年)

1986 - 2006 年为集成发展期，这是人工智能技术从理论迈向应用、从单一技术走向多学科融合的重要阶段。在这一时期，人工智能领域有了重要的技术突破且开展了应用实践，多学科与多技术的融合促使 AI 进一步发展。这些技术突破和应用实践不仅推动了人工智能自身的发展，还为深度学习以及人工智能的广泛应用奠定了基础。

1. 神经网络的复兴

1982 年，Hopfield 模型被提出，这是一种递归神经网络，能够解决一些优化问题。

1984 年，J. Hopfield 设计了 Hopfield 网络的电路，成功解决了旅行商问题（TSP），引起了广泛关注。

1986 年，Rumelhart 和 Hinton 提出了多层感知机和反向传播（BP）学习算法，克服了感知器在处理非线性分类问题上的局限，推动了深度学习的发展。

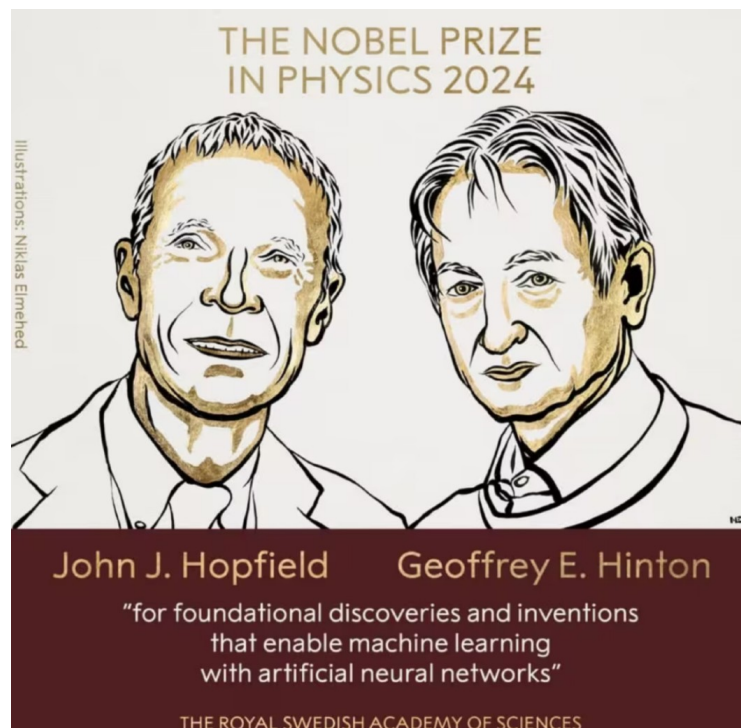


图 4 2024 年诺贝尔物理学奖得主

2. 支持向量机（SVM）

1995 年，Vapnik 提出了支持向量机（SVM），这是一种强大的分类技术，通过寻找最优超平面来区分不同类别的数据，具有良好的泛化能力，对机器学习领域产生了深远影响。

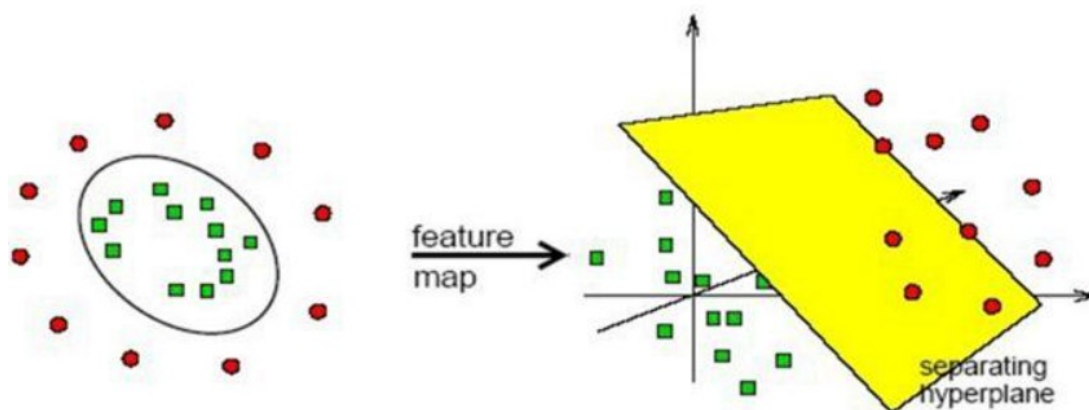


图 5 支持向量机的原理

3. 其它重要事件

1997 年 5 月 11 日,IBM 的超级计算机“深蓝”首次击败了国际象棋特级大师卡斯帕罗夫,标志着 AI 在特定领域的计算能力和策略规划上达到了新的高度。



图 6 卡斯帕罗夫与“深蓝”对决

2000 年,中国科学院计算所开发了知识发现系统 MSMiner,这是一个多策略知识发现平台,能够提供高效的数据挖掘解决方案,支持多种知识发现方法。

2011 年,IBM 的超级电脑“沃森”在美国智力竞赛节目《危险边缘》中战胜了两位最成功的选手,展示了自然语言处理和知识问答的先进能力。



图 7 IBM 的超级电脑“沃森”在美国智力竞赛节目《危险边缘》中

4. 研究领域的进展

机器学习、计算智能、人工神经网络、模糊逻辑、粗糙集理论、多智能体系统和行为主义等研究领域得到了深入的发展，推动了人工智能技术的全面进步。人工智能研究的三大学派形成

- **符号主义**：强调知识表示和推理。
- **连接主义**：关注神经网络和学习。
- **行为主义**：侧重于通过与环境的交互来实现智能行为。

这三大学派各有优势和局限，但它们的相互结合和取长补短，为人工智能的综合发展提供了坚实的基础。

1.3.7 深度学习异军突起(2006 年以来)

2006 年，Hinton 及其团队提出深度置信网络（DBN）和限制性玻尔兹曼机（RBM），在手写字符识别方面成效显著，这一成果标志着深度学习的崛起。传统神经网络在复杂结构下存在过拟合和训练困难的问题，而这些新算法和模型的出现解决了此类问题。

深度学习有诸多训练技巧，例如参数初始化方法、新型激活函数（如 ReLU）以及 Dropout 等，这些技巧能有效提升模型性能。另外，计算机和互联网的发展使大量数据的积累成为可能，从而为深度学习模型的训练提供了充足资源。

2012 年，AlexNet 在 ImageNet 竞赛中取得重大突破，其错误率从传统方法的 26.2% 降至 15.3%。AlexNet 包含 8 层网络，其中有 5 个卷积层和 3 个全连接层，参数量达 6000 万，神经元数量为 850,000 个，这表明深度学习模型的规模与计算能力正在迅速提升。

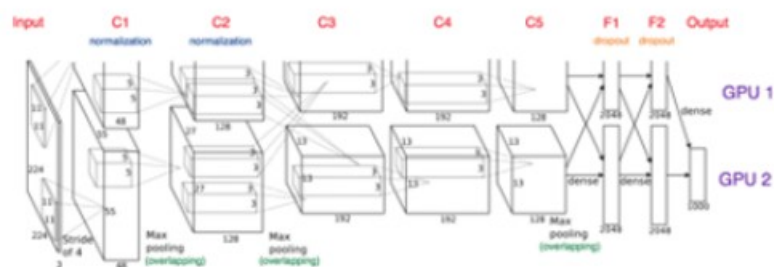


图 8 AlexNet 的网络结构

2015 年，何凯明和孙剑所在的 MSRA 团队提出 ResNet，它有 152 层，错误率达到 5.1%，超过人类识别水平。ResNet 引入残差学习，解决了深层网络训练中的梯度消失问题，使得训练更深的网络成为可能。

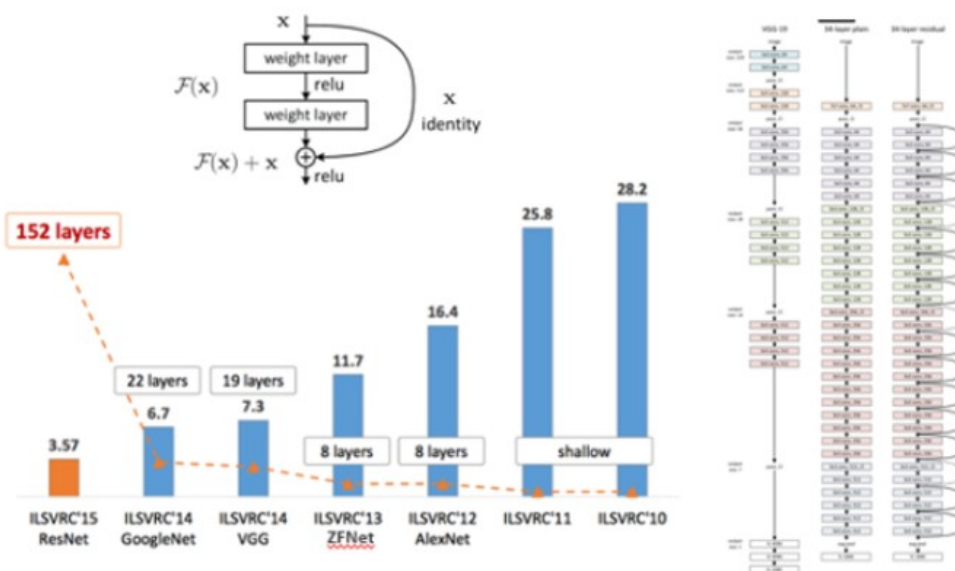


图 9 ResNet 的网络结构

1.3.8 大模型时代降临(2017 年以来)

在 2017 年，Google 公司提出了一种革命性的神经网络结构，即基于自注意力机制的 Transformer 架构。这一架构的提出，为大规模模型的预训练算法架构奠定了坚实的基础，从而开启了人工智能领域的一个新时代。

紧接着在 2018 年，OpenAI 研究团队基于 Google 的 Transformer 架构，成功发布了他们的第一个大型模型——GPT-1。这一模型的发布，标志着预训练大模型成为了自然语言处理领域的主流技术。GPT，即 Generative Pre-Trained Transformer 的缩写，是一种基于互联网海量数据进行训练的深度学习模型，它能够进行文本生成，为自然语言处理领域带来了新的可能性。

到了 2019 年，OpenAI 继续推进这一领域的发展，发布了 GPT-2 模型，进一步提升了模型的性能和应用范围。

时间来到 2020 年 6 月，OpenAI 公司推出了迄今为止最大的语言模型——GPT-3。该模型的性能达到了惊人的 1750 亿，这使得它在零样本学习任务上实现了巨大的性能提升。此外，为了进一步提高模型的推理能力和任务泛化能力，OpenAI 还引入了更多策略，如基于人类反馈的强化学习（RLHF，Reinforcement Learning from Human Feedback）、代码预训练、指令微调等。

2022 年 11 月，搭载了 GPT3.5 的 ChatGPT（Chat Generative Pre-trained Transformer）横空出世，凭借其逼真的自然语言交互与多场景内容生成能力，迅速在全球范围内引起轰动。ChatGPT 是一种由人工智能技术驱动的自然语言处理工具，它能够通过理解和学习人类的语言来进行对话，并根据聊天的上下文进行互动，真正像人类一样进行聊天交流。它甚至能够完成撰写邮件、视频脚本、文案、翻译、代码，以及撰写论文等多样化任务。

在 2023 年 3 月，OpenAI 发布了 GPT-4，这是一个多模态大模型，它不仅可以处理文本输入，还能接受图像输入，并生成文本。与上一代的 GPT-3 相比，GPT-4 在解决复杂问题方面表现得更为出色，拥有更广泛的常识和解决问题的能力。到了 2023 年 12 月，谷歌公司发布了大模型 Gemini，它能够同时识别和处理文本、图像、音频、视频和代码五种类型的信息，并且能够理解和生成高质量的主流编程语言代码，如 Python、Java、C++ 等，同时具备全面的安全性评估。

2024 年 12 月，DeepSeek 的迅速崛起震撼了全球，使得人工智能技术进入了“普惠”时代，为更多人带来了人工智能的便利。

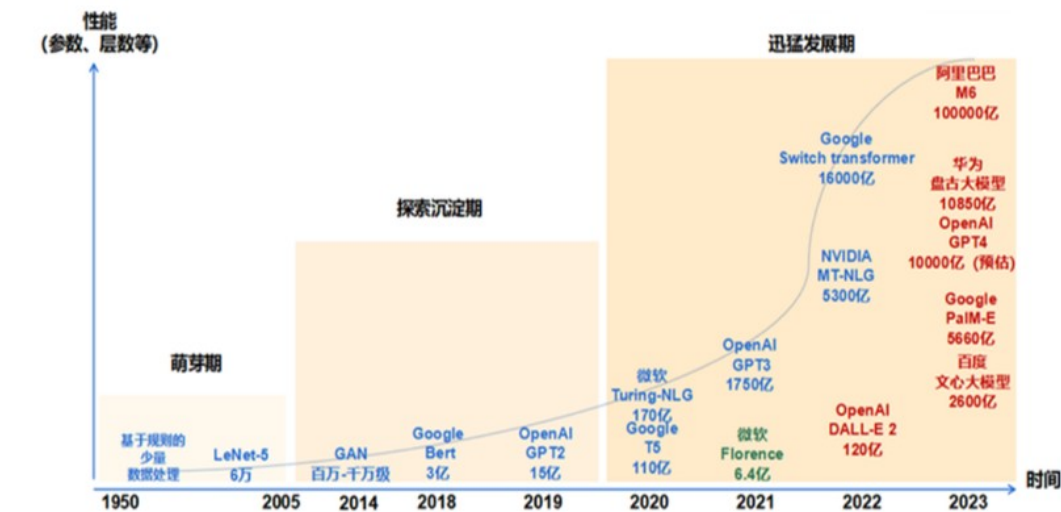


图 10 大模型的发展阶段

1.4 人工智能的研究目标

人工智能的研究是一个逐步推进的过程，从实现近期目标开始，逐步向远期目标迈进。人工智能研究的目标可以分为远期目标和近期目标。

远期目标：是构建能够实现人类思维和智力功能的智能系统，这意味着未来的 AI 应该像人一样思考、学习、解决问题，甚至具备创造力和情感理解能力。

近期目标：主要是让现有的计算机系统更聪明、更有用，能够处理数值和非数值信息，如图像识别、自然语言处理等，同时能够运用知识解决问题，模拟人类的智能行为。

远期目标和近期目标的关系：远期目标为近期目标指明方向，而近期目标为远期目标奠定理论和技术基础。通过实现近期目标，可以逐步积累知识和技术，最终实现远期目标。

OpenAI 的 5 级 AGI 量表：

OpenAI，提出了一个衡量 AI 智能水平的 5 级 AGI 量表。这个量表帮助人们理解 AI 的发展水平，并预测其未来趋势。该量表将 AI 智能分为五个等级：反应式机器、有限记忆、理论推理、自我意识和超智能，具体内容如下：

- L5：可以完成整个组织工作的 AI，具备完全自主、处理复杂任务的能力。
- L4：可以创造新事物的 AI，具备创新和发明的能力。
- L3：可以代表用户采取行动的 AI，能够理解用户需求并主动执行任务。
- L2：像博士一样解决问题的 AI，具备高级的分析和解决问题的能力。
- L1：当前的 AI，如 ChatGPT，能够与人类对话，但还不能完全模拟人类的智能。

5 级 AGI 量表为 AI 研究和发展提供清晰框架，指导未来 AI 技术的发展方向。

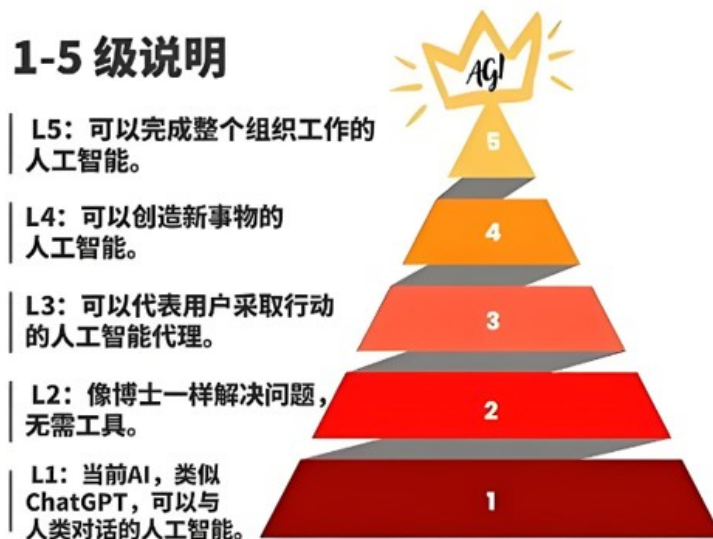


图 11 OpenAI 的五级 AGI 量表

1.5 人工智能的研究领域

目前,人工智能的研究及应用领域有很多,主要研究领域有问题求解、专家系统、智能检测、机器感知与机器学习、机器人学、分布式人工智能与多智能体等。

1. 问题求解

问题求解,即对于预定目标,寻找行动序列的过程被称为搜索,搜索算法的输入是问题,输出是问题的解,以行动序列的形式返回问题的解。搜索(尤其是状态空间搜索)和问题归纳,已成为问题求解的一种十分重要而又非常有效的手段,也是人工智能研究中的一个重要方面。目前有代表性的问题求解程序是下棋程序,计算机下棋程序涉及围棋、中国象棋、国际象棋、跳棋等,目前已达到国际锦标赛水平。另一个问题求解程序是把各种数学公式符号汇编在一起,其性能可以达到很高的水平,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。问题求解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,在人工智能中叫作问题表示的选择。人们常常能够找到某种思考问题的方法从而使求解变得容易而解决该问题。到目前为止,人工智能程序已经能够思考如何去解决问题,即搜索解答空间,去寻求较优的解答。

2. 专家系统

专家系统是一个具有大量专门知识和经验的计算机程序系统,主要由“知识库”和“推理机”组成。其根据某领域一个或多个人类专家提供的知识和经验建立知识库,应用人工智能技术和计算机技术来进行推理和判断,模拟人类专家的决策过程,解决特定领域的相关问题。简而言之,人们通过人机接口向专家系统输入需要解决的问题的相关信息,专家系统将运用推理机构控制其内在的知识库,像人类专家一样给出问题的解决方案。专家系统具有丰富的专门知识并可以模拟相关领域专家的思维过程,以解决该领域中需要专家才能解决的复杂问题。

专家系统的一般特性有:

- (1) 为解决特定领域的具体问题,除需要具备一些公共的常识,还需要具备大量与所研究领域问题密切相关的知识。
- (2) 一般采用启发式的解题方法。
- (3) 在解题过程中除了用演绎方法外,有时还要求助于归纳方法和抽象方法。
- (4) 需处理问题的模糊性、不确定性和不完全性。

- (5) 能对自身的工作过程进行推理（自推理或解释）。
- (6) 采用基于知识的问题求解方法。
- (7) 知识库与推理机分离。

专家系统可以解决的问题一般包括解释、预测、诊断、设计、规划、监控、指导和控制等。专家系统在当前以及未来都将会是人类值得信赖的高水平智能助手，是将人工智能技术运用到实际中的重要手段。高性能的专家系统也已经从学术研究开始进入实际应用研究。

3. 智能检测

智能检测技术是将传感技术、人工智能技术与智能推理技术相结合来用于参数检测的一种新型技术。随着仿生技术和模式识别等人工智能技术的发展，智能检测仪器和智能检测技术在人工智能产业中的应用得到迅速推广，其中电子鼻、电子舌等智能检测仪器在我国有着广泛的应用，电子鼻、电子舌是通过模仿人类嗅觉与味觉感官模式对物品中含有的特征物质进行辨识，利用模式识别和机器学习相关算法完成对相关物品的鉴定与分类，可以在食品安全、爆炸物品和违禁物品检测、重大疾病排查等方面广泛应用。除此之外，在工业生产过程中，运用智能检测设备与物联网技术，合理构建智能检测系统与平台，也是人工智能技术在智能检测产业广泛应用的重要表现。

4. 机器感知与机器学习

机器感知（Machine Perception）是通过解释传感器的响应为机器提供相关的信息。机器感知研究如何用机器或计算机模拟、延伸和扩展人的感知或认知能力。人工智能有各种可用的感知形态，包括类似于人类的视觉、听觉和触觉。机器学习（Machine Learning）是一种能够赋予机器学习的能力，以此让它完成直接编程无法完成的功能的方法。学习能力无疑是人工智能研究中最突出和最重要的一个方面，学习是人工智能的主要标志和获取知识的基本手段。要使机器像人一样拥有知识、具有智慧，就必须使机器拥有获得知识的能力。

使机器获得知识的方法一般有两种：

- (1) 把有关知识归纳、整理在一起，并用计算机可接受、处理的方式输入到计算机中。
- (2) 使计算机自身具有学习能力，它可以直接向书本、教师学习，也可以在实践中不断总结经验、吸取教训，实现自我不断完善，这种方式一般称为机器学习。

机器学习研究的目标有三个：人类学习机理的研究、学习方法的研究、建立面向具体任务的学习系统。机器学习与其他领域的处理技术的结合，形成了计算机视觉、语音识别、自

然语言处理等交叉学科。因此，平常所说的机器学习应用是通用的，不仅仅局限于结构化数据、还有图像、音频等应用。

5. 机器人学

人工智能研究日益受到重视的另一个分支是机器人学，机器人通过对物质世界进行操作来执行任务。目前，绝大部分机器人都属于以下三类中的一类。

(1) 操纵类机器人（机械臂）

尽管已经建立了一些比较复杂的机器人系统，不过现在工业运行的成千上万台机器人，都是一些按预定编制好的程序执行某些重复作业的简单装置。程序的生成及装入有两种方式：一种是由人根据工作流程编制程序并将它输入到机器人的存储器中；另一种是“示教-再现”方式，所谓示教是指在机器人第一次执行任务之前，由人引导机器人执行操作，即教机器人去做应做的工作，机器人将其所有动作一步步记录下来，并将每步表示为一条指令，示教结束后机器人再执行这些指令（即再现），以同样的方法和步骤完成同样的工作。

(2) 移动类机器人

该类机器人利用轮子、腿或其他机械装置，辅之以位置、避障、距离等传感器，在设定的环境中来回移动，可从事搬运、投递或信息采集等任务。该类机器人通过传感器获取作业环境、操作对象的简单信息，然后由计算机对获得的信息进行分析、处理，从而控制机器人的动作，典型代表有无人车辆、无人机等。

(3) 移动操纵类机器人

该类机器人是指具有类似于人的智能的智能机器人，该类机器人具有感知环境的能力、配备有视觉、听觉、触觉、嗅觉等感觉器官，能从外部环境中获取有关信息，它具有思维能力，能对感知的信息进行处理，以控制自己的行为，它还具有作用于环境的行为能力，能通过传动机构使自己的“手”“脚”等肢体行动起来，正确灵巧地执行思维机构下达的命令。

6. 分布式人工智能与多智能体

分布式人工智能（Distributed Artificial Intelligence, DAI）是分布式计算与人工智能结合的结果。分布式人工智能系统以健壮性作为控制系统质量的标准，并具有互操作性，即不同的异构系统在快速变化的环境中具有交换信息和协同工作的能力。分布式人工智能的研究目标是要创建一种描述自然系统和社会系统的模型。DAI中的智能并非独立存在，只能在团体协作中实现，因而其主要研究问题是各智能体（Agent）之间的合作与对话，包括分布式问题求解和多智能体系统（Multi-Agent System, MAS）两个领域。分布式问题求解把一个具体的求解问题划分为多个相互合作和知识共享的模块或者节点；多智能体系统则研

究各智能体之间行为的协调。这两个研究领域都要研究知识、资源和控制的划分问题。但分布式问题求解往往含有一个全局的概念模型、问题和成功标准。而 MAS 则含有多个局部的概念模型、问题和成功标准，MAS 更能够体现人类的社会智能，具有更大的灵活性和适应性，更适合开放和动态的世界环境，因此成为人工智能领域的研究热点。

1.6 人工智能的主要研究方法

在人工智能的研究中，主要采用了三种方法：结构模拟法、功能模拟法和行为模拟法。每种方法都有其独特的视角和研究重点，同时也存在一些局限性。

1. 结构模拟法

定义：从生物学和科学的角度出发，通过模拟人脑的生理结构和工作机理来实现人工智能。属于连接主义学派，主要研究神经网络。

不足之处：

- 人脑的结构和工作原理尚未完全理解，只能进行局部或近似模拟。
- 不适合模拟逻辑思维过程。
- 受到大规模人工神经网络制造的限制。
- 无法完全模拟人脑。

2. 功能模拟法

定义：从认知学和工程的角度出发，通过符号主义学派的方法，将问题或知识表示为逻辑结构，运用符号演算来实现表示、推理和学习等功能，从而在宏观上模拟人脑思维，实现人工智能功能。

不足之处：

- 符号表示的有效性取决于符号表示的正确性和准确性。
- 在将知识概念转换成符号时，可能会丢失重要信息。
- 难以处理含噪信息、不确定性信息和不完全性信息。

3. 行为模拟法

定义：从行为主义学派的角度出发，认为智能取决于感知和行动，智能行为只能在现实世界中与周围环境交互作用而表现出来。

不足之处：

难以获得高级智能控制行为。

1.7 关于课程

本课程的目标是全面而深入地向学习者展示人工智能领域的丰富多样性和其独特的特点，确保学习者能够对人工智能这一领域有一个全面而深刻的理解。通过本课程的学习，学习者将能够掌握人工智能的核心概念和原理，从而为他们未来在相关领域的深入研究和专业课程的学习奠定坚实的基础。因此，课程内容广泛而全面，不仅包括了知识表示与推理、计算智能、机器学习、深度学习、大模型技术等人工智能的核心领域，还特别强调了具身智能等前沿研究方向，以期让学习者能够紧跟人工智能技术的最新发展和趋势。