1 绪论

1.1 人工智能

1.1.1 智能的概念

智能的本质一直是哲学与脑科学领域的重要研究课题，且被列为自然界四大未解之谜（物质本源、宇宙诞生、生命本质及智能起源）。尽管近年来学界对人脑结构与机能形成了基础性认知，但其工作原理与运作机制仍存在大量未知。这导致目前仍无法从根本上对智能形成明确界定。现阶段研究多基于已有知识体系与实证经验，通过多维度视角与方法融合阐释智能内涵。其中比较有代表性的理论有思维理论、知识阈值理论及进化理论等。

（1）思维理论源于认知科学领域。该理论强调智能的本质在于思维过程，认为人类的知识活动与认知行为均根植于大脑的思维运作，所有知识皆为思维系统生成的成果。因此，通过探索思维规律及其逻辑推演的运行机理，可为揭示智能本质提供思路。

（2）知识阈值理论的核心观点聚焦知识在智能中的价值与功能，主张智能表现由知识储备量及其泛化水平决定，系统具备智能的根本前提是拥有知识体系。将智能界定为在庞杂解空间中高效定位可行解的效能。该理论深刻影响了人工智能发展进程，知识建模、专家决策系统等技术均在其理论框架下得以发展。

（3）进化理论由美国麻省理工学院（MIT）布鲁克教授于1992年提出并倡导。其“无推理智能”的核心理念主张以动态控制替代符号表征，消解概念模型与显性知识表达的必要性，弱化抽象思维在智能模拟中的核心地位。

综合上述观点，可将智能视为知识体系与认知能力的综合产物。其中，知识构成所有智能行为的基础，而认知能力体现为知识获取及其应用的问题求解能力，即在任意给定的环境和目标条件下，正确制订计划和实现目标的能力。

智能应具备以下特征。

（1）感知功能：作为外部信息采集的基本途径。

（2）记忆存储与思维能力：记忆模块负责存储感知系统捕获的外部数据及思维生成的知识体系；思维单元构成动态处理机制，通过知识获取与应用实现问题的求解。

（3）学习能力和自适应能力：通过持续学习完成知识演进，实现对外部环境变迁的响应优化。

（4）行为能力：改造世界的能力，可用作信息的输出。

1.1.2 人工智能

简言之，人工智能(artificial intelligence，AI)就是让机器具有人类的智能，这也是人类科技发展长期探索的方向。人脑历经数亿年进化形成了复杂结构，但我们至今仍然没有完全了解其工作机理。虽然随着神经科学、认知心理学等学科的发展，人们对大脑的结构有了一定程度的了解，但对大脑的智能究竟是如何产生的还知之甚少。我们尚未完全了解其运行机制，尤其是意识、情感及记忆等功能的形成机理，这使现阶段实现人工智能仍面临理论瓶颈。

1950年，计算机科学先驱艾伦·图灵发表了具有里程碑意义的论文《计算机器与智能》，探讨了“智能机器”的可行性，并提出了划时代的图灵测试：测试者通过隔离交互界面进行连续问答，若在预设时限内无法辨别对话方是人类还是计算机，则可判定该计算机具备智能。通过该测试需整合语言理解、知识学习、逻辑推理及自主决策等核心模块，这促使人工智能研究从思辨层面转向实证科学，并衍生出感知计算(视觉解析、语音处理)、认知计算(模式识别、机器学习)、语义计算(自然语言处理)、知识图谱构建及决策系统(智能规划、数据挖掘)等关键技术体系。

1956年的达特茅斯会议是人工智能学科体系化的关键节点，会上正式确立了“人工智能”这一学科名称并明确其研究使命。约翰·麦卡锡在此次会议提出的学科界定中强调：人工智能致力于构建能模拟人类智能行为特征的机器系统。

目前，人工智能的主要领域大体上可以分为以下几个方面。

(1) 感知。模拟人的感知能力，对外部刺激信息(视觉和语音等)进行感知和加工。主要研究领域包括语音信息处理和计算机视觉等。

(2) 学习。模拟人的学习能力，主要研究如何从样例或从与环境的交互中进行学习。主要研究领域包括监督学习、无监督学习和强化学习等。

(3) 认知。模拟人的认知能力。主要研究领域包括知识表示、自然语言理解、推理、规划、决策等。

1.2 人工智能的发展历程

人工智能的发展历程大致可以分为推理期、知识期和学习期三个阶段。

1.2.1 推理期

1956年达特茅斯会议后，人工智能领域迎来研究热潮，其后十余年进入技术高速发展期。早期学者主要依据人类经验，先通过逻辑推演与事实归纳构建规则，再借助算法编码实现计算机的特定任务。该阶段诞生了几何定理自动证明系统、跨语言转换模型等技术，这些突破性进展导致学术界对实现通用人工智能产生过度预期，对技术实现的复杂性存在系统性低估。当时普遍存在“20年内机器将全面替代人类劳动”“3～8年可研制出具有人类平均智能的机器”等激进论断。随着认知科学研究的深化，学术界逐渐发现，推理模型的局限性和项目难度的不可控性使原有技术路线遭遇理论瓶颈，直接导致研究投入锐减与学术信任危机。人工智能研究发展滞缓期(亦称“研究寒冬”)是指该领域经历技术成熟度曲线的峰值

后，因理论突破不足引发的资金撤离与学术关注度下降现象，其中 1974—1980 年与 1987—1993 年是两段具有代表性的低谷时期。

1.2.2 知识期

20 世纪 70 年代，人工智能研究发生认知转向，学术界系统认识到领域知识在智能系统中的核心地位。尤其是针对复杂问题求解场景，亟须构建专家级知识架构体系。此阶段涌现出各种类型的专家系统，并在专业领域实现技术突破。可将专家系统简单理解为“知识库＋推理机”，它是一类具有专门知识和经验的计算机智能程序系统，通过知识表征方法与推理逻辑算法，复现领域专家的决策方式，故此类系统被界定为基于知识的系统。标准专家系统需满足三大建构准则：①领域专家级知识；②模拟专家思维；③达到专家级水平。在此阶段，Prolog 逻辑编程语言成为核心开发工具，其作为一种面向演绎推理的逻辑型程序设计语言，特别适用于构建专家决策系统、智能知识库以及处理自然语言理解。

1.2.3 学习期

在此阶段，科研人员逐渐将研究焦点转移至计算机自主学习能力。早在人工智能发展初期，已有学者致力于构建机器的自我学习机制，即机器学习（machine learning，ML）。该技术通过构建和解析学习算法，使计算系统能够从数据中自主分析并归纳规律，继而运用其规律对未知信息进行推断，最终协助人类完成专项任务。机器学习的研究范畴广泛，涉及线性代数、概率论、统计学与数值计算等诸多基础学科。尽管机器学习早期已是人工智能的核心分支之一，但直至 20 世纪 80 年代后才逐步发展为备受瞩目的研究热点。

1.3 人工智能的研究方法

目前我们对人类智能的机理依然知之不多，还不能基于一个通用的理论构建人工智能系统。不同的研究者对其理解不同，因而在人工智能的研究过程中便产生了许多不同的观点。一种观点主张基于计算机科学的方法进行研究，实现人类智能在计算机上的模拟；另一种观点主张用生物学的方法进行研究，搞清楚人类智能的本质。前一种方法是以符号处理为核心的方法，又称符号主义；后一种方法采用以网络连接为主的连接机制，又称联结主义。除此之外，还有一种基于“感知—行动”的行为智能模拟方法，该方法主张人工智能起源于控制论，称为行为主义。

1.3.1 符号主义

符号主义（亦称逻辑主义、心理学派或计算学派）起源于 20 世纪 50 年代中期，由学者纽厄尔、西蒙等提出，旨在复现人类问题求解的认知机制，并由此构建物理符号系统理论。该学派主张人工智能的核心目标在于实现机器智能，强调计算机的符号运算机制天然具备逻辑推演特性，这种特性本质映射了形式逻辑的推理内涵，因而能够通过特定程序架构展现逻辑导向的智能表征，以实现对人类认知活动的数字化模拟。因其理论体系以符号操作为基础，故称为符号主义。符号主义的主要特征如下：

(1) 擅长复现人类逻辑推理路径,适用于需形式化推演的复杂问题。

(2) 知识可用显式的符号表示,在系统先验规则明确时,无须输入大量的细节知识。

(3) 具备模块化架构优势,面对局部信息更新时系统维护更便捷。

(4) 可实现与传统符号数据库的无缝对接与数据交互。

(5) 对结论具有可解释性,便于对各种可能性进行选择。

需要注意的是,求解问题时非逻辑推理往往起着更重要的作用,甚至是决定性作用。同时,人的感知过程主要是形象思维,这也是逻辑推理所不能及的。此外,用符号表示概念时,其有效性在很大程度上取决于符号表示的正确性,当把有关信息转换为推理机构能进行处理的符号时,将丢失一些重要信息。这些都是制约符号主义解决智能问题能力的因素。

1.3.2 联结主义

联结主义也称仿生学派或生理学派,是在人脑神经元相互联结构成网络的前提下,试图通过人工神经元间的并行协同作用实现对人类智能的模拟。该方法认为大脑是人类一切智能活动的基础,应对大脑神经元及其联结机制进行研究,搞清楚大脑的结构以及进行信息处理的过程与机理,从而真正实现人类智能在机器上的模拟。联结主义的主要特征如下:

(1) 依托神经单元并行协作机制完成信息运算,具备并行计算、动态演化和整体关联特性。

(2) 采用分布式物理连接结构存储知识信息,支持模式联想功能,对噪声、残缺及形变数据具备容错处理能力。

(3) 通过神经联结权值的自适应调节机制,实现人类认知行为的模拟。

(4) 擅长模拟人类非逻辑的形象思维过程。

(5) 问题求解过程中能快速生成满足精度要求的近似解。

1.3.3 行为主义

行为主义的核心体现为感知加行动,强调在复杂的真实物理场景中进行认知训练,相比符号主义基于形式逻辑、联结主义依托数据驱动,其环境适应训练面临更大困难。该理论体系源于控制论。控制论思想在 20 世纪 40—50 年代已成为现代科学的重要组成部分。维纳与麦卡洛克等提出的控制论及自组织理论体系、钱学森团队发展的工程控制论与仿生控制理论等,深刻影响了多学科发展进程。本质上,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了起来。早期研究聚焦于模拟人类智能调控行为,重点探索自寻优、自适应、自组织、自学习等智能调控机制,并开展仿生控制论实体研制。20 世纪 60—70 年代,相关研究取得阶段性突破,不仅为智能机器人技术奠定了理论基础,更推动了 20 世纪 80 年代智能控制系统与智能机器人系统的诞生。

1.4 人工智能算法

自 1956 年达特茅斯会议提出“人工智能”的概念后,人工智能便作为一门学科正式进入研究者视野,与之相关的研究和算法层出不穷。人工智能算法经过几十年的发展,从专家系统问世掀起人工智能算法研究的高潮,到 21 世纪深度神经网络的再一次蓬勃发展,其间涌

现出众多在各自领域成果斐然的算法。从广义上看，人工智能算法包括神经网络及深度学习、强化学习、群智能算法、搜索算法、不确定性推理、监督学习、无监督学习及半监督学习等方法。本书中人工智能算法的主要组成如图 1-1 所示。

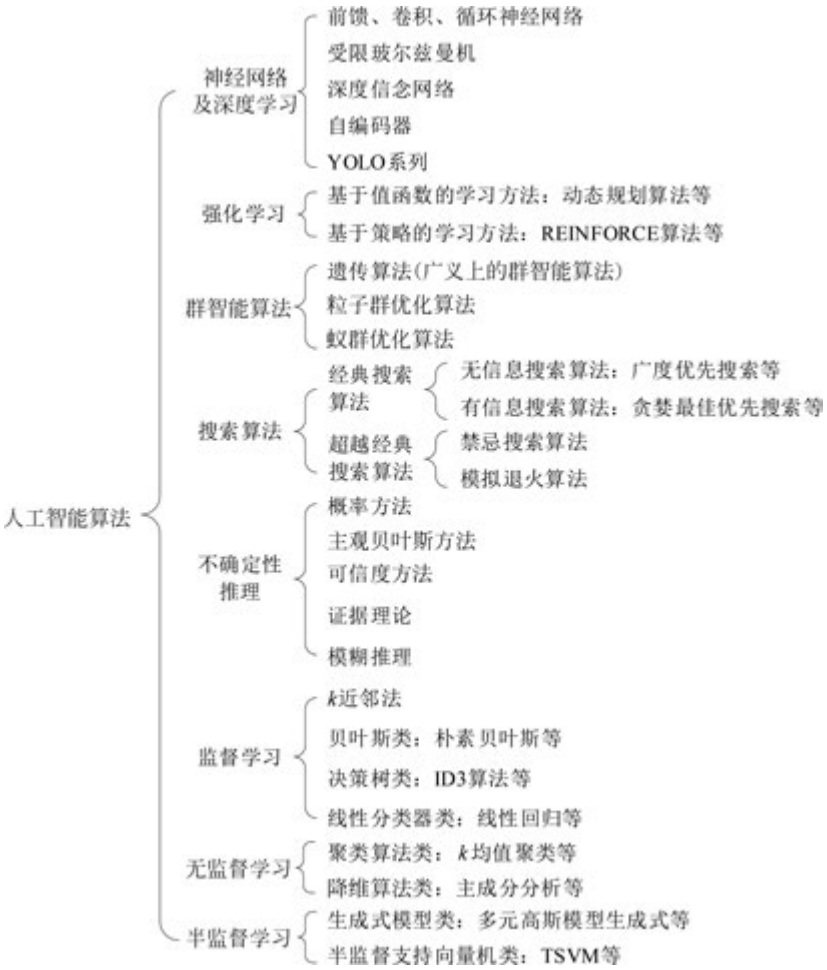

图 1-1 人工智能算法的主要组成

1.4.1 神经网络及深度学习

神经网络受生物神经元启发，通过多层节点连接模拟人脑的信息处理机制。深度学习作为其进阶形态，通过构建深层网络结构，利用海量数据实现自动化的特征提取与抽象。它克服了传统机器学习依赖人工设计特征的局限，极大地提升了算法在复杂环境下的泛化能力，是当前人工智能领域最核心的驱动技术。神经网络示意图如图 1-2 所示。

常见神经网络以及深度学习模型如下。

(1) 神经网络：前馈神经网络(feedforward neural network，FNN)、卷积神经网络(convolutional neural network，CNN)和循环神经网络(recurrent neural network，RNN)等。

(2) 深度学习模型：受限玻尔兹曼机(restricted boltzmann machine，RBM)、深度信念

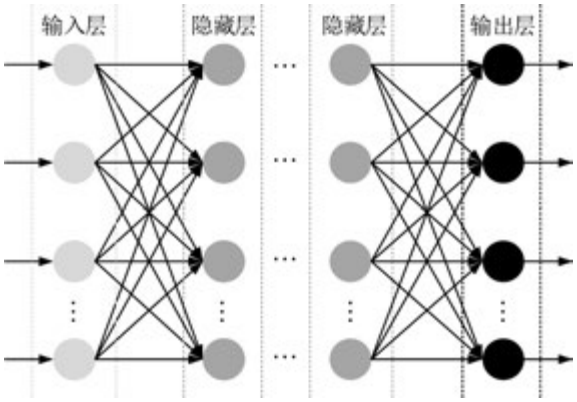

图 1-2 神经网络示意图

网络(deep belief network,DBN)、自编码器(autoencoder,AE)以及 YOLO(you only look once)系列等。

1.4.2 强化学习

如图 1-3 所示,强化学习是通过智能模型不断与环境进行交互,基于试错的方式进行学习,以获得最佳策略。相比于监督学习,强化学习没有大量的数据和标签,需要从与环境的交互中得到这些数据。

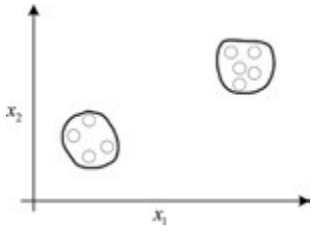

图 1-3 强化学习

强化学习的主要算法如下。

(1) 基于值函数的学习方法:动态规划算法、蒙特卡罗强化学习和时序差分学习等。

(2) 基于策略的学习方法:REINFORCE 算法和带基准线的 REINFORCE 算法等。

1.4.3 群智能算法

群智能(swarm intelligence,SI)算法是借鉴蚂蚁、鸟群等社会性动物的群体智能行为研究得出的算法,是人们从生物群体活动的自然现象中发现并加以探究的结果。常见的群智能算法有粒子群优化算法、蚁群优化算法等。遗传算法是受生物群体行为启发,并通过多个简单个体的相互作用,在系统层面具有智能解决问题能力的算法,从这个层面来看,遗传算法可以认为是广义上的群智能算法。因此,本书将遗传算法纳入第 5 章群智能算法中进行介绍。

1.4.4 搜索算法

搜索算法是利用计算机的计算性能有目的地穷举某种问题的部分或者全部解空间,从而得到该问题的解。在大量实际工程应用中,通常通过降低搜索规模、根据约束剪枝等方法对算法进行优化。搜索算法主要分为经典搜索算法和超越经典搜索算法。

经典搜索算法又分为无信息搜索算法和有信息搜索算法。无信息搜索算法主要包括广度优先搜索、一致代价搜索和深度优先搜索等。有信息搜索算法主要包括 A-star 算法等。

超越经典搜索算法主要包括禁忌搜索算法和模拟退火算法等。

1.4.5 不确定性推理

不确定性推理是解决现实世界复杂性与不可预测性的核心手段。鉴于客观数据的随机性、模糊性或不完整性，传统的精确逻辑往往难以奏效。该技术通过引入概率论、模糊数学及证据理论等工具，使智能系统能够在知识不完备或存在噪声的条件下，进行合理的推断并量化结论的可信度。它赋予了算法在动态环境中进行鲁棒决策的能力，是模拟人类柔性思维的关键环节。常见的方法包括概率方法、主观贝叶斯(Bayes)方法、可信度方法、证据理论以及模糊推理等。

1.4.6 监督学习

监督学习的实施流程如下：构建标注完备的数据集集合，将其导入预设的参数化模型结构，通过迭代优化使模型性能收敛。训练过程中系统将持续比对模型输出与样本真实标签的差异，直至预测精度达到预设阈值标准。监督学习常用于解决分类(如图 1-4 所示)和回归问题。

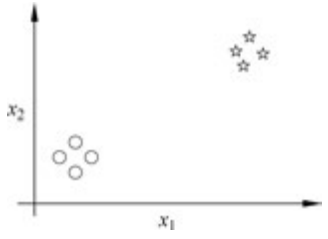

图 1-4 监督学习用于解决分类问题

常见监督学习算法如下。

(1) k 近邻法(k-nearest neighbors，KNN)。

(2) 贝叶斯：朴素贝叶斯(naive Bayes，NB)和贝叶斯网络(Bayesian network，BN)等。

(3) 决策树：分类和回归树(classification and regression tree，CART)、ID3 算法、C4.5 算法和 C5.0 算法等。

(4) 线性分类器：线性回归、逻辑回归和支持向量机等。

1.4.7 无监督学习

如图 1-5 所示，无监督学习方式下，输入数据并没有明确的标识，将数据输入算法模型，模型通过使用大量的数据推断出数据的内在结构。

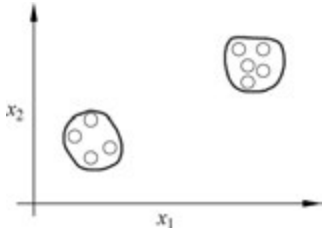

图 1-5 无监督学习

常见无监督学习如下。

(1) 人工神经网络：生成对抗网络(generative adversarial network，GAN)和前馈神经网络(feedforward neural network，FNN)等。

(2) 关联规则学习：先验算法(apriori algorithm，AA)等。

(3) 聚类算法：通常依据划分、层次、密度、网格、模型、图等不同策略进行分类，常用的有层次聚类和 k 均值聚类等。

(4) 降维算法：常见的有主成分分析(principle component analysis，PCA)和奇异值分解(singular value decomposition，SVD)等。

1.4.8 半监督学习

如图 1-6 所示，半监督学习是一种由监督学习与无监督学习相结合的学习方法，也是监督学习的一种延伸。该方法使用大量的标记数据和无标记数据，先对未标识数据建模，再对标识数据进行预测。

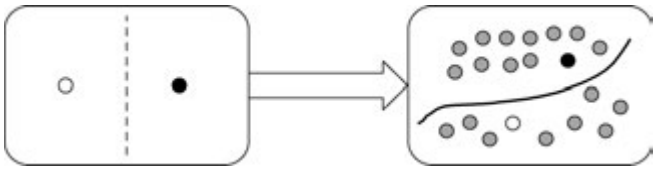

图 1-6 半监督学习

常见的半监督学习算法包括生成模型(generative model)、低密度分离(low-density separation)和基于图形的方法(graph-based method)等。

1.5 人工智能的研究领域

1.5.1 专家系统

专家系统是人工智能中活跃且有成效的一个研究领域。自费根鲍姆等研发出第一个专家系统 DENDRAL 以来，它已获得快速发展，广泛应用于医疗诊断、地质勘探、石油化工、教学、军事等各领域，获得了巨大的社会效益和经济效益。

专家系统是一种基于知识的系统，从人类专家处获得知识，解决只有专家才能解决的困难问题。可以这样定义专家系统：专家系统是一种具有特定领域内大量知识与经验的程序系统，模拟人类专家求解问题的思维过程解决领域内的各种问题，其水平可以达到甚至超越人类专家。

1.5.2 机器学习

作为人工智能的一个研究领域，机器学习主要研究如何使计算机具有类似人类的学习能力，使计算机通过学习自动地获取知识与技能，实现自我完善。研究内容主要包括三个方面：人类学习机理的研究、学习方法的研究和建立面向具体任务的学习系统。

机器学习与脑科学、神经心理学、计算机视觉、计算机听觉等都有密切联系，并依赖这些学科的共同发展。目前已经取得很大的研究进展，提出了多种学习方法，如之前介绍的监督学习、无监督学习、半监督学习、强化学习等方法。

1.5.3 模式识别

机器感知是机器智能的一个重要方面，是机器获取外部信息的基本途径。模式识别是研究如何使机器具有感知能力的重要研究领域，主要研究对视觉模式和听觉模式的识别。

模式是对一个物体(或实体)定量的或者结构化的描述，而模式类是指具有某些共同属性的模式集合。模式识别的主要内容是研究一种自动技术，使机器能自动地把模式分配到各自的

模式类中。传统的模式识别方法主要包括统计模式识别和结构模式识别两大类。目前模糊数学、人工神经网络等技术已经应用于模式识别，出现了模糊模式识别、神经网络模式识别等方法。

1.5.4 自然语言处理

自然语言处理的研究源于突破传统人机交互壁垒。早期计算设备依赖专业化编程实现功能调度，这种技术门槛导致计算机应用长期局限于专业群体，极大地制约了信息技术的推广。若计算系统具备语音识别与视觉解析能力，则将显著提升人机交互功能，进而大大提高计算设备的利用率。自然语言处理便是研究如何使计算机理解人类自然语言的一个领域，研究内容主要包括自然语言理解和自然语言生成两个方面。

1.5.5 自动定理证明

自动定理证明是人工智能早期核心研究领域之一，其理论突破与实践验证为知识表示、推理算法等关键技术的发展提供了支撑。定理证明的实质在于，对前提 P 和结论 Q，证明 P→Q 的永真性。但是，直接证明 P→Q 的永真性一般来说比较困难，通常采用反证法。在这方面，海伯伦(Herbrand)与鲁宾逊(Robinson)先后进行了卓有成效的研究，提出了相应的理论与方法，为自动定理证明奠定了理论基础。尤其是鲁宾逊提出的归结原理，使自动定理证明得以在计算机上实现，对机器推理做出了重要贡献。

1.5.6 自动程序设计

自动程序设计包括程序综合与程序正确性验证两个方面。程序综合用于实现自动编程，即用户只需向计算机明确任务目标，无须指导具体实现步骤，计算机就可自动实现程序设计；程序正确性验证是指研究出一套理论和方法，运用这套理论和方法即可证明程序的正确性。目前常用的验证方法是用一组已知结果的数据对程序进行测试，如果程序的运行结果与已知结果一致，就认为程序是正确的。这种方法对于简单程序来说容易实现，但对于复杂系统来说就比较困难。因为复杂程序中存在纵横交错的关系，具有难以计数的通路，很难保证程序的正确性。程序正确性验证至今仍是一个比较困难的课题，还需进一步研究。

1.5.7 机器人学

人工智能的所有技术几乎都可应用于机器人，机器人可用作人工智能理论、方法、技术的试验场地。同时，对机器人学的相关研究又推动了人工智能研究的发展。总的来说，机器人学的研究已经经历了从低级到高级的三代发展历程。

1. 程序控制机器人

第一代机器人是程序控制机器人，是完全按照程序安排的步骤进行工作。程序的生成与装入有两种方式：一种是由人根据工作流程编制程序，并将其输入机器人的存储器；另一种是示教再现型，即在机器人第一次执行任务之前，由人引导机器人执行操作，机器人将其所有动作一步步记录下来，并将每一步表示为一条指令，示教结束后机器人通过执行这些指令以同样的方式和步骤完成同样的工作。

2. 自适应机器人

相比第一代机器人，第二代机器人自身配备感觉传感器，如视觉传感器、触觉传感器、听觉传感器等，并通过计算机进行控制。机器人通过传感器获取外部信息，由计算机对获得的信息进行分析、处理，并控制机器人的行为。由于这种机器人能随着环境的变化而改变自己的行为，故称为自适应机器人。目前，这一代机器人已进入商品化阶段，主要从事焊接、装配、搬运等工作。第二代机器人虽然具有一些初级的智能，但还没有达到完全自治的程度。

3. 智能机器人

智能机器人是具有类似人的智能的机器，即具有环境感知能力，配备视觉、听觉、触觉、嗅觉等感觉传感器，能从外部环境中获取有关信息；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；具有环境改造能力，能通过传动机构使自己的手、脚等肢体行动起来，准确地执行思维机构下达的指令。

1.5.8 博弈

人工智能研究博弈的目的是将博弈环境作为实验平台，检验人工智能技术能否实现对人类智能的模拟。博弈研究通过提供复杂的决策场景，推动了算法的优化与理论突破。1956年，塞缪尔研制出了跳棋程序；1991年8月，第12届国际人工智能联合会议上，IBM的Deep Thought与澳大利亚国际象棋冠军约翰森(D. Johansen)进行了人机对抗赛，最终比分为0∶2(人类胜)；1996—1997年，“深蓝”与卡斯帕罗夫进行了两次对决，首次以2∶4落败，次年以3.5∶2.5获胜；在国际公认难度最高的围棋领域，2016年3月Google的AlphaGo以4∶1战胜韩国棋手李世石，成为首个在无让子条件下击败人类职业九段选手的AI系统。

1.5.9 智能决策支持系统

智能决策支持系统(intelligent desision support system，IDSS)是20世纪80年代中期形成的交叉领域，其核心由专家系统等人工智能技术与传统决策支持系统融合而成。21世纪，社会进入了一个数据爆炸式增长、环境复杂多变的时代，人类的大多数决策都以大量的数据信息为基础，超出了人类自身的信息处理能力。IDSS通过四库(数据库、模型库、方法库、知识库)系统，将定量分析的决策支持系统与定性分析的专家系统结合，显著提高了辅助决策能力。

习题

1. 什么是人工智能？试从学科和能力两方面加以说明。
2. 在人工智能发展史中，哪些思想和思潮起到了重要作用？
3. 21世纪以来，人工智能的发展方向有哪些？
4. 为什么能够用机器模仿人的智能？
5. 现代人工智能有哪些学派？其认知观是什么？
6. 人工智能发展史中的基本研究方法分为几类？各有何应用？
7. 如今的大数据时代，人工智能涌现出哪些新的研究特点？为什么？