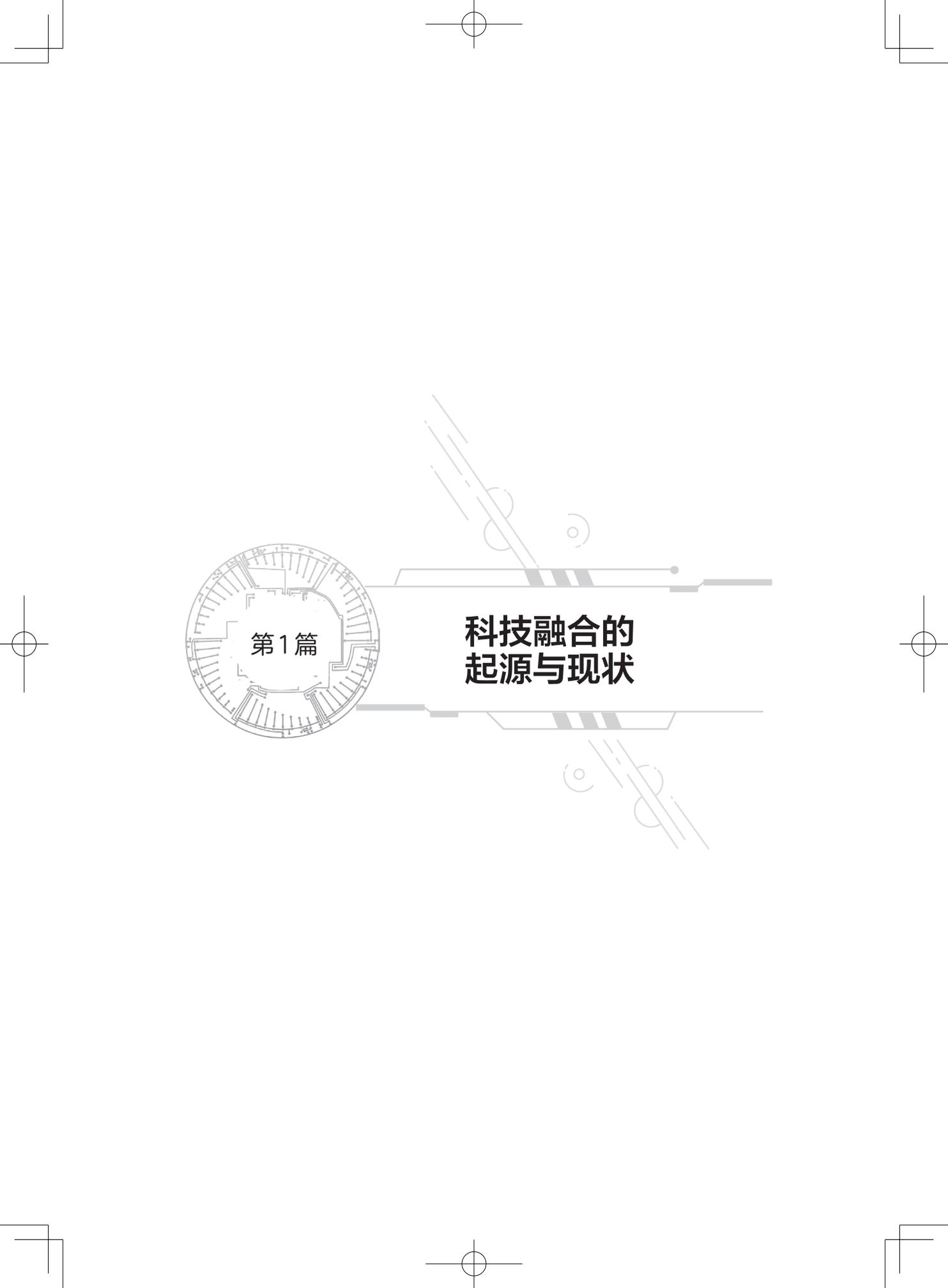


科技融合的 起源与现状



引言

AI与机器人技术如何从独立领域走向融合

2025年，具身智能（Embodied Intelligence）技术正加速突破物理与数字世界的边界，推动机器人从“机械执行”向“自主决策”跃迁。作为杭州“六小龙”的核心成员，宇树科技与深度求索（DeepSeek）的协同创新，成为这一浪潮的典型缩影。

2025年央视春晚上，宇树科技人形机器人H1以《秧BOT》表演惊艳全球（如图 I.1 所示），其技术突破与应用潜力标志着具身智能走进日常生活的临界点已至。该系列机器人搭载先进 AI 驱动全身运动控制技术，能精准复刻传统秧歌的韵律与神韵，头部 3D 激光雷达（LiDAR）和 360° 全景感知系统使它们如同真人般规避碰撞、感知环境。表演中故意设计的“失误彩蛋”更展现出科技与人文趣味融合的创新思维，预示机器人未来将具备更强的交互亲和力。



图 I.1 宇树科技人形机器人H1以《秧BOT》表演惊艳全球

从实际应用场景看，宇树 H1 身高 180cm、体重 47kg 的类人体型设计，已初步适配家庭服务场景需求。随着规模化生产推动成本下降，预计 3 ~ 5 年内其售价将降至万元以下，成为中等收入家庭的“智能管家”，承担端茶倒水、物品搬运、作业辅导等任务。而技术迭代将加速功能拓展——通过持续学习用户习惯，未来机器人不仅能执行指令，还能主动预判需求，例如在主人回家前调节室温、准备餐食。

这一趋势背后是多方合力的推动：政府通过政策引导加速科技创新与文化产业融合，企业的技术突破带来机器人运动控制、环境交互等核心能力的质变，市场对智能终端的旺盛需求则催生新经济增长点。正如春晚舞台将科技符号转化为文化记忆，具身智能的普及也将重塑生活方式，使人类从重复劳动中解放，从而能够专注于更具创造力的领域。当机器人能流畅完成东北秧歌的“转手绢”动作时，它们提菜篮、擦窗户的日常化服务，自然也就触手可及了。

宇树科技凭借全球领先的四足机器人技术，已实现工业巡检、应急救援等场景的规模化应用，其最新一代机器狗搭载 DeepSeek 自研的多模态大模型，可自主识别复杂环境，动态规划路径，并在化工园区巡检中实现故障预判准确率超 98%。而 DeepSeek 作为通往国产通用人工智能（Artificial General Intelligence, AGI）的领军者，其大模型技术不仅为机器人赋予“大脑”，更在医疗、教育等领域推动具身智能与行业需求的深度融合。宇树科技联合能源巨头打造的“AI 巡检员”，通过多传感器融合与实时数据分析，将传统人工巡检效率提升 10 倍以上，单台设备年降本超百万元。

浙江省以杭州“六小龙”为支点，整合机器人硬件、AI 算法与行业数据资源，形成了全球首个具身智能全栈技术链，吸引微软、特斯拉等国际巨头争相共建联合实验室。

未来 3 年，具身智能将呈现两大核心趋势。

技术融合深化：硬件（如柔性关节、仿生感知）与软件（如因果推理、多模态交互）的协同创新，将催生“感知—决策—执行”一体化的下一代机器人。

场景指数级扩展：从工业制造向医疗护理、农业种植、太空探索等长尾场景渗透，预计 2030 年全球市场规模将突破万亿美元。

人工智能（AI）与机器人技术的融合是一个快速发展的领域，近年来在国际上受到了广泛的关注。这种融合不仅推动了技术的进步，也为多个行业带来了革命性的变革。

波士顿动力公司开发的 Atlas 人形机器人在 2024 年 4 月展示了其最新版本。这款机器人具有 28 个电动关节，能够在复杂地形上行走，并且能够举起和操纵负载。Atlas 的设计考虑了工业应用，尽管尚未广泛应用于汽车工厂，但考虑到波士顿动力现在归现代汽车所有，这种应用可能很快就会实现。

具身智能的最终实现，或者说完美的具身智能，必须先要实现空间智能（Spatial Intelligence）。现有的尝试只是用大语言模型来作为具身物理体的大脑，但这肯定不完美，毕竟语言和文字只是展示一维的信息，并不是真正的理解物理世界本身，初期效果只是期望达到物理感知和语言大模型间的融合。而空间智能是实现了理解三维物理世界的智能，是把感知以及与物理世界互动所产生的新鲜数据再训练和建模，从而形成感知、推理、决策、行动的链条，这可能是通往最终的通用人工智能的路径。而人形机器人是具身智能的一种典型的应用形式。

美国斯坦福大学的李飞飞团队认为，空间智能是指机器在三维空间和时间中感知、推理和行动的能力。了解物体在三维空间和时间中的位置，以及与物理世界的互动如何影响其三维位置。三维位置在时空中交互，将机器从主机或数据中心中带到现实的三维世界中来，使其更好地理解三维和四维世界。

AI 技术，尤其是大模型，能够处理和分析海量数据，提供精准的决策支持。当这些能力与机器人的物理执行功能结合时，可以极大地扩展机器人在复杂环境下的应用范围，提升其自主性和适应性。例如，服务机器人在酒店服务、高级制造、医疗辅助等领域的应用，通过 AI 技术的提升，能够更精准地理解任务并生成控制代码，减少了对人工编程的依赖。

特斯拉的人形机器人 Optimus 在 2023 年 12 月展示了其第二代版本，具有更先进的执行器和传感器。Optimus 在特斯拉工厂中进行了测试，能够执行精确的电池分类任务。它展示了在低容错率环境下的工作能力，并能自主修正错误。

根据特斯拉发布的视频，Optimus 利用端到端神经网络技术，通过 2D 摄像头、手部触觉和力传感器收集数据，实时在 FSD 计算机上运行，以实现精确的电池单元分拣和插入托盘的任务。在这一过程中，Optimus 能够自动定位下一个空闲槽位，并且在出现错误时，能够自主地从故障中恢复，调整动作以完成任务。

Optimus 的这种自主性是通过大量的数据训练实现的。特斯拉通过人类远程操作机器人来收集数据，并对各种任务进行扩展训练。随着时间的推移，Optimus 在工厂测试中的人工干预率持续下降，表明其自主性能正在不断提升。此外，Optimus 的手部设计也是其能够进行精细操作的关键，其手部拥有 22 个自由度，能够进一步提高其操作的灵活性和精确度。

特斯拉 CEO 埃隆·马斯克强调，Optimus 已经在执行工厂任务，并预计将在未来几年内更广泛地使用。Optimus 的这些进步展示了人形机器人在工业自动化和智能制造中的巨大潜力，以及 AI 技术在机器人领域应用的快速发展。

大模型，特别是大语言模型在机器人任务级交互中发挥着关键作用。这些模型不仅包含大量的参数和高计算资源需求，还能够处理复杂的任务并取得卓越的性能。例如，谷歌 DeepMind 的研究者利用深度学习来加快发现新材料的过程，所用技术被称为“材料探索图网络 (GNoME)”，已经预测了 220 万种新材料的结构，其中 700 多种已经在实验室中实现了合成。

特斯拉的 Optimus 人形机器人被设计用于执行传统上需要人工干预的任务，其应用范围不仅限于电池分类。Optimus 的先进神经网络和特斯拉专有的 AI 技术使其能够处理复杂任务并适应多样化的操作环境。类似的一些潜

在工业自动化任务包括：

（1）物体识别与分类：Optimus 能够识别和分类物体，如在工厂中对不同颜色的块进行分类并放入相应的托盘中。它甚至能够在动态变化的环境中继续执行分类任务，例如在人类工人移动块后继续进行分类。

（2）物流与仓储：Optimus 的自主导航能力使其能够在仓库环境中导航，从而优化仓库操作并减少人为错误。

（3）装配线任务：Optimus 的设计使其能够模仿人类动作，这使得它能够执行需要精细操作的任务，如组装线上的部件装配。

（4）材料处理：Optimus 可以处理重复性的材料搬运任务，减少人工劳动并提高效率。

（5）螺丝拧紧和组装：Optimus 可以执行螺丝拧紧和其他紧固应用，这些通常是工业产品中重复性高且易于自动化的任务。

（6）人形机器人与人类的协作：Optimus 被设计为能够在人类旁边安全工作，这为制造自动化提供了新的可能性，尤其是在装配领域。

（7）灵活的制造自动化：Optimus 可以推动制造自动化从传统的刚性自动化向适应性强、有弹性的自动化转变。

（8）模拟和数字孪生：通过模拟和数字孪生技术，Optimus 可以在实际部署前在虚拟环境中进行测试和验证，以优化其性能和安全性。

Appttronik 公司开发的 Apollo 机器人专为仓库作业设计，具有模块化设计，可以安装在不同的平台上。德国汽车制造商梅赛德斯—奔驰计划在其工厂测试 Apollo 的腿部版本，用于物流和装配套件交付。

Figure AI 公司制造的 02 机器人在宝马位于南卡罗来纳州的工厂中参与汽车组装，负责将金属片零件插入固定装置并组装成底盘的一部分。机器人领域的任务级交互指的是机器人能够从接收具体任务指令到完成具体动作的全过程中的自主操作。这种交互模式显著提高了机器人的操作效率和适用范围，因为它减少了人类操作者的介入，使得机器人能够在更广泛的环境和情境下独立作业。

相关研究进展表明，AI 与机器人技术的结合正在不断深化。例如，斯

斯坦福大学以人为本人工智能研究所发布的《2024年人工智能指数报告》中提到，产业界继续主导人工智能前沿研究，并且人工智能在某些任务上胜过人类，但并非在所有任务上。此外，报告还指出，中国在人工智能专利数量上处于遥遥领先的地位，这表明中国在AI与机器人技术融合方面的研究和应用具有显著的全球影响力。

NASA开发的人形机器人Valkyrie在2023年7月被部署在澳大利亚的石油和天然气生产设施中，展示了其在非结构化环境中的工作能力。

AI与机器人技术的融合对经济产生了显著影响。例如，麦肯锡的调查显示，42%的受访组织报告部署人工智能技术降低了成本，59%的组织报告收入实现了增加。这表明AI技术正在推动业务效率的大幅提高。

随着AI与机器人技术的融合，相关的政策和法规也在不断发展。例如，美国的人工智能相关法规数量急剧增加，这表明政府正在积极参与，鼓励人工智能的发展，同时也在管理潜在的不利因素。

AI技术，尤其是大模型，正在与机器人技术融合，提供更精准的决策支持和自主性。这种融合在工业自动化、智能制造、医疗辅助等领域的应用正在增加。机器人技术正在从传统的工业应用扩展到更广泛的领域，如农业、医疗保健、物流和服务业。这些应用不仅提高了效率，还有助于解决劳动力短缺问题。

中国政府在2023年11月发布了《人形机器人创新发展指导意见》，提出到2025年初步建立人形机器人创新体系，并在特种、制造、民生服务等场景实现批量生产和示范应用。AI与机器人技术的融合正在推动技术进步和行业变革，从工业自动化到服务机器人的智能化，这种融合正在开启智能自动化的黄金时代。随着技术的不断发展和应用的广泛推广，我们可以期待未来会有更多的可能性和变革出现。

目前，国际和国内TOP5的人形机器人公司的主要发展路线各有侧重，通过不断的技术创新和应用场景探索，推动了人形机器人产业的发展。

例如，特斯拉的Optimus机器人在设计上借鉴了人类生物学特征，如膝关节骨骼和五指的生理结构，以提高其灵活性和操作精度。Optimus的手部

具身智能与空间智能：人形机器人应用

是最先进的五指灵巧机器人手之一，具有触觉感应功能，能够感知和处理各种物体。这种设计使得 Optimus 在执行任务时更加独立和有效，能够处理更复杂的任务，如电池分类和物品搬运。

特斯拉的 AI 技术在 Optimus 中的应用还包括使用特斯拉的全自动驾驶（FSD）控制器，这为机器人提供了高级的视觉处理能力和实时决策制定功能。通过这些技术，Optimus 能够在没有人类直接监督的情况下，自主完成复杂任务，如在工厂环境中移动电池和在家庭环境中处理待洗衣物等。随着技术的不断进步，Optimus 在未来的工业自动化和智能制造中将发挥越来越重要的作用。

特斯拉计划将其首先应用于汽车工厂，执行移动搬运、零部件装配等工业级操作，随后可能扩展至家庭服务等更复杂的环境，使其成为通用服务机器人。特斯拉的 Optimus 预计将在这两年量产，并在特斯拉工厂进行实用性测试。特斯拉的优势在于其可以直接嫁接 FSD 感知与算法、深度神经网络训练软硬件基础，且具备顶级的研发团队和成熟的汽车供应链。

还有老牌机器人公司波士顿动力（Boston Dynamics），波士顿动力以其卓越的运动控制和动态性能而闻名，其 Atlas 机器人展示了强大的平衡能力和越障能力。波士顿动力的人形机器人在运动性能上处于行业领先地位，其技术沉淀和多样化的下游应用使其成为该领域的主要玩家之一。

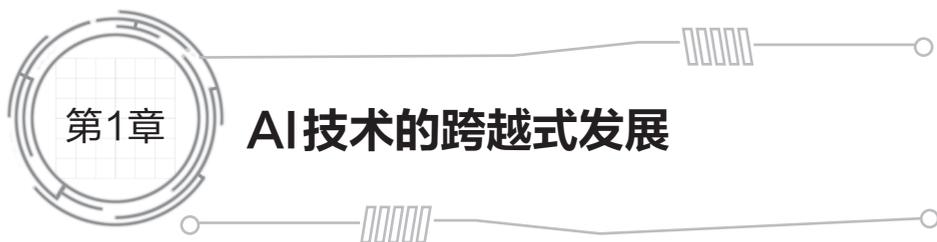
在大模型领衔的赛道上，IX Technologies 的人形机器人 NEO 采用了“无齿轮”设计理念，全身较为柔软，能够在几乎所有场景中安全作业。NEO 结合了先进的语言系统和智能自主运动模型，能够在与人类互动中自主学习，展示了具身智能的潜力。

国内机器人企业如智元机器人、优必选等，它们通常拥有一定的机器人技术积累，并在近年来开始发力人形机器人赛道。智元机器人的远征 A1 预计先应用在工业制造领域，随后逐步走向 ToC 应用。优必选的 Walker X 主要应用于科技展馆、影视综艺、商演活动、政企展厅等场景。

同时，AI 与机器人交汇贯通的过程中，互联网企业也杀出了一条血路，以小米、科大讯飞为代表的国内互联网企业也是人形机器人赛道的潜在竞争

者。小米发布了 CyberOne，专注于重建真实世界、实现运动姿态平衡、感知人类情绪。这些企业通常拥有强大的资金实力和算法技术，通过设立事业部或子公司的方式切入人形机器人赛道。

各公司在人形机器人领域的发展路线主要集中在技术创新、应用场景探索，以及与现有业务的协同效应上。随着技术的不断进步和市场需求的不断增长，人形机器人产业有望在未来几年内实现更广泛的商业化应用。



第1章 AI技术的跨越式发展

1.1 早期专家系统的局限性与突破

1.1.1 早期专家系统的局限性

AI技术的跨越式发展在人形机器人领域的体现是显著的。从早期的专家系统开始，AI技术经历了从局限到突破的过程。早期的专家系统虽然在特定领域内能够模拟专家的决策能力，但它们通常缺乏灵活性和自适应能力，难以处理复杂和多变的现实世界问题。

比如，DENDRAL——一个早期的化学专家系统，用于帮助识别有机化合物的结构。尽管DENDRAL在特定领域内表现出色，但它的知识和推理能力仅限于化学分析，难以扩展到其他领域。再比如，MYCIN，这是一个早期的医疗专家系统，用于辅助诊断细菌感染并推荐治疗所需抗生素。MYCIN系统的主要局限在于它只能处理有限的几种疾病，且其决策树结构难以适应复杂多变的医疗情况。

随着时间的推移，专家系统开始采用机器学习方法来自动化知识获取过程。例如，Query by Analogy系统能够通过学习新案例来扩展其知识库，减少了对专家的依赖。

为了解决早期专家系统缺乏常识推理的问题，研究者开发了如CYC项目，它试图构建一个包含广泛常识知识的数据库，使AI系统能够更好地理解和推理现实世界的问题。

早期专家系统通常采用确定性推理，难以处理不确定性信息。Probabilistic Reasoning 系统如 BAYESS (Bayesian Expert System Simulator) 通过引入概率论来处理不确定性和模糊性，提高了系统的鲁棒性。

随着深度学习的发展，专家系统开始融合多种数据模态。例如，Watson (IBM 开发的问答系统) 不仅能够理解自然语言，还能处理视觉和听觉数据，提供了更为丰富的交互体验。

现代 AI 系统如 AlphaGo 和 AlphaZero 通过深度学习和强化学习，能够自我对弈和学习，不断优化其策略和决策能力，这在早期专家系统中是难以实现的。

通过这些例子可以看出，早期专家系统的局限性在于它们通常只能处理特定领域的狭窄问题，缺乏灵活性和自适应能力。而随着 AI 技术的发展，特别是在机器学习、深度学习、多模态学习和强化学习等领域的突破，现代 AI 系统已经能够处理更加复杂和多变的问题，展现出更高的智能水平。

早期专家系统的局限性主要体现在以下几个方面：

(1) 知识获取难题：早期专家系统依赖于专家知识，但知识的获取、表达和更新是一个复杂且耗时的过程。

(2) 缺乏常识推理：这些系统通常缺乏对现实世界常识的理解，导致在面对未预见情况时表现不佳。

(3) 难以处理不确定性：早期系统往往难以处理不确定性和模糊性，这限制了它们在动态环境中的应用。

在人工智能的发展历史中，技术进步是连续且有序的。其中，有几个重要的发展阶段，它们显著地推动了人工智能领域的发展。早期的人工智能系统主要依赖于人类输入的明确指令和知识。机器学习技术的出现，使得这些系统能够从大量数据中自动学习和提取模式。例如，IBM 的 Watson 系统能够自主地分析大量信息，并提供基于数据的答案，这标志着人工智能在自动学习和知识获取方面的一大进步。

1.1.2 早期专家系统的突破与深度学习的发展

深度学习技术的发展进一步扩展了人工智能的能力，特别是在图像和语音识别、自然语言处理等领域。它使得人工智能系统能够识别复杂的模式和细微的差别。例如，谷歌 DeepMind 的 AlphaGo 通过深度学习技术，不仅学会了围棋的策略，还在对弈中战胜了人类顶尖选手，这显示了人工智能在策略和决策方面的显著进步。

随着多模态学习技术的发展，人工智能系统开始整合多种类型的数据，如视觉、听觉和文本信息，以获得更全面的理解。例如，微软的 Cognitive Services 能够分析图像内容并生成描述，这表明人工智能在综合不同信息源和提供丰富输出方面的能力正在增强。

空间智能与具身智能的交互不仅体现在技术实施层面，更在于它们带来的思维方式的转变。随着智能系统的普及，我们开始重新审视人机协作的模式，强调人类与机器之间的互补性而非替代性。这一转变不仅影响了技术的设计，也改变了我们对工作的理解。

在未来的工作环境中，智能体将成为人类的合作伙伴，而不是简单的工具。通过增强的空间智能，这些智能体能够理解人类的意图，并根据环境变化作出快速反应。这一协作模式能够有效提升生产效率，减少人为错误。在教育领域，空间智能和具身智能的结合为学习方式带来了革命性的变化。通过虚拟现实（VR）和增强现实（AR）技术，学生可以在模拟环境中进行实践操作，如医学培训、工程设计等。这种沉浸式学习方式不仅提高了学习效果，也加深了学生对知识的理解。

在智能城市的建设中，空间智能和具身智能同样发挥着重要作用。通过部署无人机和自动化监测设备，这些技术可以实时收集城市环境数据，如空间地理信息、交通流量、空气质量等，帮助城市管理者制定科学的决策。这种基于数据的决策过程，凸显了智能体在复杂城市环境中对空间智能的需求。

这些技术的发展不仅是人工智能领域的技术突破，也是人类对智能本质理解的深化。

1.2 现代深度学习的革命性应用案例

1.2.1 深度学习在自动驾驶领域的应用

现代深度学习技术的发展已经带来了多个革命性的应用程序，这些应用正在改变我们与技术的互动方式，并让我们在多个行业中实现创新。

例如，在自动驾驶汽车领域，深度学习正在推动自动驾驶技术的发展。通过处理来自摄像头、传感器和地图的数据，深度学习模型能够训练车辆在交通中导航，识别路径和交通标志，并实时响应交通状况。例如，Uber的人工智能实验室正在使用深度学习来提升无人驾驶汽车的性能。

自动驾驶汽车的发展是深度学习技术应用的一个重要领域。深度学习通过模拟人脑神经元的工作方式，使得自动驾驶汽车能够处理大量数据，识别复杂的道路环境，实现环境感知、决策规划和控制执行等关键功能。

自动驾驶汽车利用深度学习算法，如卷积神经网络（CNN），来识别道路标志、交通信号灯、行人和其他车辆。这些算法能够从车载摄像头获取的图像中提取特征，提供车辆周围环境的详细视图。例如，特斯拉的自动驾驶系统通过深度学习模型，实现了从感知到控制的无缝连接，极大地提升了自动驾驶的效率和安全性。

在感知环境后，自动驾驶汽车需要作出决策，如何时加速、减速或转向。深度学习技术可以帮助汽车学习并模拟人类的驾驶行为，实现智能决策规划。深度强化学习（Deep Reinforcement Learning, DRL）通过试错学习，使自动驾驶汽车在不断尝试和反馈中优化驾驶策略。

自动驾驶汽车在作出决策后，需要执行相应的操作。深度学习技术可以帮助实现对车辆动力、制动和转向系统的精确控制。例如，基于深度学习的控制器可以学习并预测车辆在不同路况下的动态响应，以实现更平稳、更安全的驾驶。

随着技术的进步，自动驾驶汽车正逐步从试验走向商用化。多家科技巨头和传统汽车制造商正在推动自动驾驶技术的商业化进程，例如特斯拉、百

具身智能与空间智能：人形机器人应用

度等企业正在进行大量的研究和测试。未来，自动驾驶技术有望构建起一个高效、安全、环保的智能交通生态系统，进一步提升出行效率和安全性。

笔者于 2024 年 12 月参访北京市高级别自动驾驶示范区，从具身智能与空间智能视角来看，其具有如下三个核心特点：

➤ 车路云一体化技术框架

示范区构建“车、路、云、网、图”五大协同体系，通过网联云控系统实现车辆与城市基础设施的实时数据交互，支持 L4 级以上自动驾驶车辆的规模化运行。其“多杆合一、多感合一”的路侧感知系统覆盖 28 个路口、110 处点位，形成全域动态感知网络。

➤ 具身智能的动态环境适配能力

自动驾驶车辆搭载多模态传感器（激光雷达、摄像头等）与自主决策算法，在复杂城市环境中实现精准导航与实时避障。例如，小马智行 Robotaxi 通过持续学习人类驾驶行为优化“老司机思维”，在动态交通场景中降低其车辆的驾驶难度。

➤ 空间智能驱动的城市级协同优化

云端智能中枢（三级云架构）整合全域交通数据，实现交通信号优化（车均延误下降 30%）、绿波带协调（道路平均速度提升 12.3%）等空间治理目标。高精度地图与数字孪生技术支撑虚拟测试环境与物理空间的映射验证。

在实际运行中，新石器、智行者等企业的无人车在南海子公园执行商品配送任务，支持 65℃ 保温餐盒与 -18℃ 冷藏雪糕的精准温控交付。

在全无人 Robotaxi 服务方面，小马智行自动驾驶车辆在亦庄全域开展常态化运营，2022 年年底实现全车无人化测试，累计自动驾驶里程超 1449 万公里。

在智慧物流调度方面，示范区部署无人配送车执行“最后一公里”运输，通过云端路径规划系统动态规避拥堵路段，日均完成 3000 多件物流订单。北汽与百度在改造停车场测试自主代客泊车功能，车辆通过 V2X 通信实时接收车位状态与路径指引，泊车效率提升达到 40%。无人环卫车与巡逻车在深

夜执行街道清扫、安防巡检任务，其行动轨迹与城市管理需求实时匹配，能够减少 70% 的人工干预。

示范区正在推动自动驾驶技术从“单车智能”向“车、路、云一体化”升级（如图 1.1 所示），通过具身智能体与空间智能系统（如全域数字孪生平台）的深度融合，加速实现交通、制造、服务等领域的全场景自动化重构。

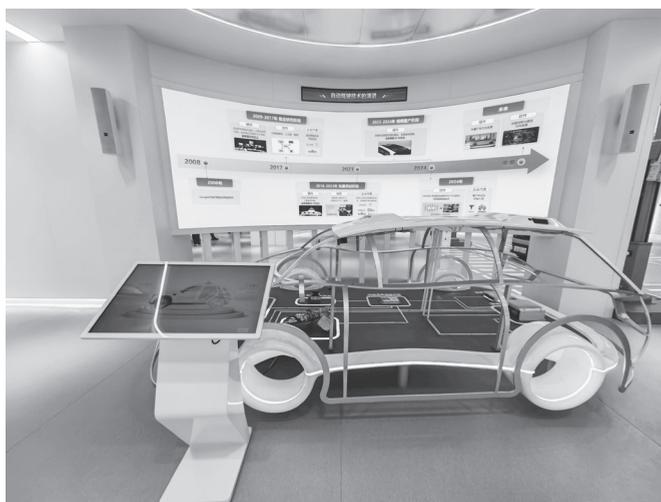


图1.1 北京市高级别自动驾驶示范区

同时，自动驾驶技术的发展也面临着技术安全性、法律法规、数据隐私保护等挑战。确保技术的可靠性和透明度，同时制定合理的监管框架，对于保护公众利益至关重要。随着自动驾驶汽车的普及，就业结构和社会习惯也将发生深刻变化，如何应对这些挑战将是未来政策制定者、行业参与者以及各利益相关方需要考虑的重要议题。

自动驾驶汽车在实际应用中遇到的技术难题依然很多。尽管自动驾驶技术取得了显著进展，但在硬件、软件和计算能力等方面的整合效果仍有待提高。例如，激光雷达穿透雨雾的能力有限，易受强光干扰；摄像头在夜间和恶劣天气中的感知灵敏度下降，这限制了自动驾驶系统在复杂环境下的性能。

自动驾驶功能的实现需要额外的成本，包括摄像头、激光雷达、V2V/V2I、

处理器等。随着自动驾驶级别的提高，对算力的需求也显著提高，这导致车辆成本增加，可能限制了自动驾驶技术的普及。

自动驾驶不仅需要车辆自身的硬件和软件发展，还需要通信端、路端、云端等基础设施的支持，北京市高级别自动驾驶示范区的自动驾驶巴士已经开始在相关区域试运行（如图 1.2 所示），这将为北京市城市级交通网络延伸示范区 4.0 阶段奠定良好的基础。



图1.2 北京市高级别自动驾驶示范区的自动驾驶巴士

未来，北京将构建“自动驾驶北京环”（四环至六环），扩展机场、高铁站等枢纽接驳，并启动京津冀高速干线物流自动驾驶测试。当前示范区累计发放 900 张测试牌照，自动驾驶巴士日均处理物流订单 3000 多件，测试总里程超 3200 万公里（占全国总量 25%），关键路口通行效率提升 20%。

自动驾驶技术是数据驱动的，需要大量且场景丰富的数据来训练和迭代算法。然而，自动驾驶场景中的数据呈现出长尾分布特征，边缘场景（corner case）出现概率低，获取这些数据需要大量行驶里程的积累，这是非常耗时且成本高昂的工作。

随着自动驾驶技术的发展，现有的法律法规在权责认定、道德伦理等方面存在较大争议和缺失。例如，当自动驾驶车辆发生事故时，责任划分和赔

偿责任等问题尚未明确，这限制了自动驾驶技术的进一步应用。

自动驾驶系统在处理复杂光照变化和极端天气条件下的表现仍有待提高。例如，在日出和日落时，自动驾驶车辆的事故率可能远高于人类驾驶，这表明自动驾驶技术在环境适应性方面还需进一步的技术突破。

不难看出，自动驾驶汽车的实际应用需要克服技术成熟度、成本、基础设施、数据丰富度和法律法规等多方面的挑战，才能实现更广泛的商业化。

1.2.2 深度学习在医疗领域的应用

在医疗领域，深度学习被用于早期诊断疾病，如癌症检测和癫痫发作预测。它还可以帮助医生通过分析医学影像来提高诊断的准确性，同时在药物发现和基因组分析中发挥作用。

例如，在癌症检测方面，深度学习算法，尤其是卷积神经网络（CNN），已被用于分析医学影像，以识别和定位肿瘤。比如，YOLOv8 算法被用于开发癌症图像检测系统，该系统能够精确识别医疗图像中的癌症标记，支持图片、视频和实时摄像头输入，并具备热力图、类别标记、统计分析等高级功能。

在癫痫发作预测方面，基于深度学习的技术，如递归神经网络（RNN）和长短期记忆网络（LSTM），已被用于分析脑电图（EEG）数据，以预测癫痫发作。这些模型能够学习时间序列数据中的特征，从而提前预警患者可能的癫痫发作。

在医学影像分析方面，深度学习模型，如 U-Net 和 EfficientNet，被用于医学影像的分割和分类任务。这些模型能够处理 CT、MRI 等成像数据，帮助医生进行更准确的诊断。

在药物发现方面，深度学习在药物发现领域的应用包括预测药物分子的生物活性和潜在毒性。通过分析大规模的生物医学数据，深度学习模型可以帮助识别潜在的药物靶点和药物分子，加速新药物的发现和研发过程。

另外，在基因组分析方面，深度学习技术也被用于分析患者的基因组数据，帮助医生识别潜在的遗传变异和风险因素，为个性化治疗提供依据。

这些应用展示了深度学习在医疗保健领域的潜力，尤其是在提高诊断准确性、预测疾病风险和加速药物研发方面。然而，这些技术的应用也面临诸多挑战，包括数据隐私和安全性、模型的可解释性以及数据多样性的处理等。

1.2.3 深度学习在计算机视觉等领域的应用

深度学习在计算机视觉领域的应用已经推动了图像识别和视频分析技术的显著进步，这些技术正在安全监控、内容过滤和社交媒体等多个方面发挥着重要作用。

在安全监控领域，深度学习技术被用于实时分析视频流，以识别异常行为或潜在威胁。例如，通过训练模型识别特定模式或行为，安全系统可以自动报警，提高响应速度和准确性。

在社交媒体上，深度学习技术被用于图像和视频的自动标记、分类和搜索。例如，Facebook 利用深度学习技术自动标记和组织用户上传的照片，使得用户能够更容易地找到和分享他们的内容。社交媒体和内容平台使用深度学习来自动过滤不当内容，如暴力、色情或其他违规内容。这些系统可以分析图像和视频内容，快速识别并处理不适宜公开展示的材料。

比如，谷歌的研究团队开发了一种名为“*In Silico Labeling*”的技术，它使用深度学习直接对细胞影像生成荧光标记，这种方法可以预测多种荧光标记，而无须对细胞进行实际的物理标记，这在生物学和医学研究中具有重要意义。

深度学习技术也在推动智能助理的发展，如 Siri、Google Assistant 等，它们通过深度学习来理解和响应自然语言命令，提供更加个性化的服务。这些智能助理不仅能够回答问题，还能执行任务，如设置提醒、播放音乐、发送消息等，极大地改善了用户体验。

深度学习技术还能不断创造更加沉浸式的 AR 和 VR 体验，通过理解用户的环境和行为来提供个性化的内容和交互。借助分析和理解用户的环境，深度学习技术能够在用户的 AR/VR 体验中精确地放置虚拟对象。例如，通过分析来自传感器的数据，AR/VR 系统可以识别并构建环境的 3D 模型，实现更智能的物体放置和交互。

AR 和 VR 还可以加强个性化内容呈现，深度学习算法可以根据用户的行为和偏好提供个性化的内容推荐。在教育 and 游戏应用中，系统可以根据用户的互动和反馈调整内容，提供定制化的体验。

比如，银河通用机器人在智能家居管理中，如图 1.3 所示，GALBOT G1 通过双臂协同与躯干伸缩能力（操作范围覆盖 0 ~ 2.4 米），可自主完成地面清扫、高处物品存取、衣物晾晒等任务。其“跪姿”模式支持稳定抓取地面杂物，“站立”模式可触及吊柜顶部，配合开柜门、抽屉等泛化技能，实现全屋空间智能管理。

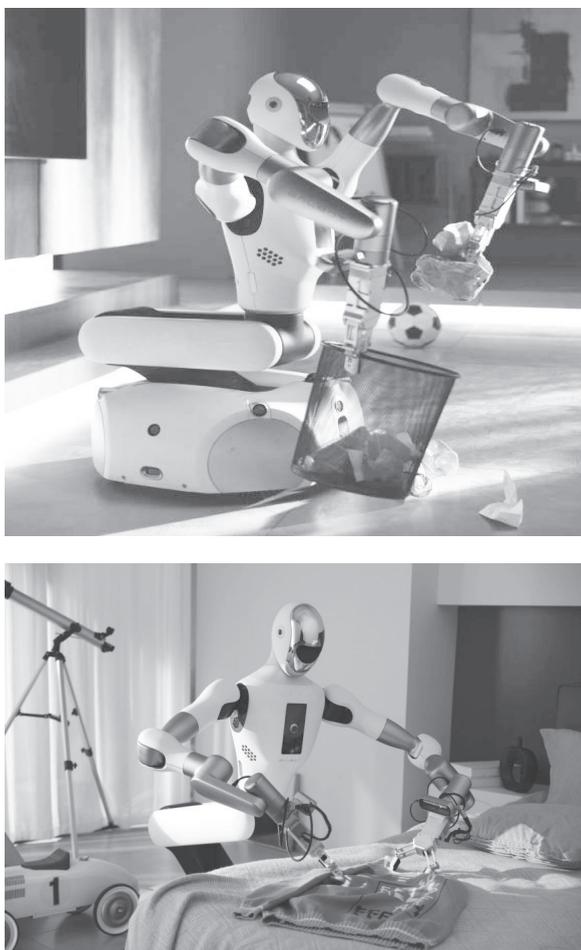


图 1.3 银河通用机器人在智能家居管理环境中工作

具身智能与空间智能：人形机器人应用

基于多模态大模型技术，机器人可理解自然语言指令，执行夜间定时巡查、药品递送、跌倒监测等任务。例如，通过 VLM(视觉语言模型) 识别儿童哭闹情绪并启动安抚程序，或为老人播放定制化健康提醒。

机器人还支持 65°C 保温餐盒与 -18°C 冷藏物品的精准存取，可以满足家庭热食保温、生鲜配送等需求。其移动底盘支持 360 度转向，可在狭小厨房空间内完成餐具整理、食材分拣等精细化操作。

在语音识别与交互方面，深度学习增强了 AR/VR 中的语音识别能力，使得用户可以通过自然语言与虚拟环境进行交互。这不仅提高了用户的沉浸感，也使得操作更加直观和友好。

由于深度学习技术的加持，互动式学习与培训成为可能的现实。在教育和专业培训领域，深度学习技术可以监测学习者的进度，并提供个性化的学习路径和即时反馈，从而提高其学习效率。

新的技术使得艺术家和创意工作者可以更好地创造出新的艺术形式和表达方式。例如，通过 AR 技术，艺术家可以在现实世界中叠加虚拟元素，创造出独特的艺术体验。

在零售和营销领域，AR/VR 技术结合深度学习为用户提供虚拟试穿、虚拟展示等服务，提升了购物体验 and 营销效果。VR 技术还能提供沉浸式的游戏体验，如《水果忍者 VR》等游戏，通过 VR 设备可提供更加真实的互动和娱乐体验。

这些应用案例展示了深度学习如何从理论走向实践，解决现实世界的问题，并在多个领域创造了新的可能性。随着技术的不断进步，我们可以期待深度学习在未来将带来更多的创新和突破。