人工智能通识（微课版）

马晓燕　编著

清華大學出版社
北　京

内 容 简 介

本书系统介绍了人工智能领域的核心理论与关键技术，全面展现了从基础概念到前沿应用的人工智能知识体系。本书共分7章，涵盖了人工智能概述、人工智能的应用、机器学习技术、深度学习技术、大语言模型概述、智能体与具身智能、AIGC应用与实践等内容。

本书结构清晰，案例丰富，与书中内容同步的教学视频可供读者随时扫码学习，读者可扫描案例旁边的二维码在线观看教学视频。本书既注重基础理论的系统性，又突出技术发展的前沿性，可作为高等院校和培训机构的人工智能相关课程的授课教材，也可供广大科技工作者和对人工智能感兴趣的读者自学使用。

本书配套的电子课件和素材文件可以通过扫描前言中的二维码获取。读者扫描正文中的二维码可以观看教学视频。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。
版权所有，侵权必究。举报：010-62782989，beiqinquan@tup.tsinghua.edu.cn。

图书在版编目(CIP)数据

人工智能通识 : 微课版 / 马晓燕编著. -- 北京 : 清华大学出版社, 2026. 5. -- ISBN 978-7-302-71380-7

Ⅰ. TP18

中国国家版本馆CIP数据核字第2026QD6147号

责任编辑：胡辰浩
封面设计：高娟妮
版式设计：妙思品位
责任校对：成凤进
责任印制：丛怀宇

出版发行：清华大学出版社
网　　址：https://www.tup.com.cn，https://www.wqxuetang.com
地　　址：北京清华大学学研大厦 A 座　　邮　　编：100084
社 总 机：010-83470000　　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印 装 者：三河市铭诚印务有限公司
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张：**14.25　　**字　　数：**338 千字
版　　次：2026 年 5 月第 1 版　　**印　　次：**2026 年 5 月第 1 次印刷
定　　价：69.80 元

产品编号：114244-01

前言

我们正处在一个由人工智能驱动的技术变革时代。从智能手机中的语音助手到改变科研范式的AlphaFold，从提升生活便利性的推荐系统到重塑知识工作方式的DeepSeek，人工智能已不再是科幻作品中遥不可及的构想，而是深刻影响社会发展、产业升级和人类生活方式的现实力量。

本书的编写初衷，是为读者提供一部系统、全面且与时俱进的人工智能通识教材。我们深知，在信息爆炸的今天，初学者往往面临知识碎片化、技术更新快、理论与实践脱节等挑战。为此，本书在内容编排上兼顾基础性与前瞻性，既深入讲解人工智能的核心原理，又充分展现其最新发展态势，力求为读者构建完整的人工智能知识体系。通过学习本书，读者可以掌握人工智能的基本概念、原理和技术，并了解其在各个领域的应用。

本书具有以下特色。

- 体系完整：涵盖从基础理论到尖端应用的完整知识链。
- 内容新颖：包含大语言模型、智能体、AIGC等最新技术进展。
- 实践导向：通过丰富的应用案例展现技术的实际价值。
- 难易适中：在保证专业性的同时，力求把知识点表述得清晰、易懂。

本书系统介绍人工智能领域的核心理论与关键技术，全面展现了从基础概念到前沿应用的知识体系。本书共分为7章，每章主要内容如下。

第1章：对人工智能的概念进行探讨，明确其技术要素与科学基础，梳理其发展历程并展望技术方向，以帮助读者构建系统性的认识。

第2章：探讨人工智能在制造、医疗、金融、教育、零售等多个行业中的具体应用场景，分析人工智能如何重塑传统行业，并开创未来发展的更多可能性。

第3章：介绍机器学习的基础框架与关键技术，涵盖机器学习的核心任务、典型方法、实现策略及基本要素，为理解与应用机器学习奠定基础。

第4章：从深度学习基本概念出发，系统阐述深度学习与机器学习的区别、发展脉络、核心模型架构及其典型科研应用场景。

第5章：介绍大语言模型的基本原理、技术架构与工程实践，探讨了规模扩展与数据工程两大核心驱动力。同时介绍了应用开发中的关键方法与主流评估体系，揭示了其从理论走向广泛应用的完整路径。

第6章：深入探讨了智能体和具身智能的基本概念及其在人工智能领域的重要性。首先阐述

了智能体的应用场景和关键特征，并介绍了智能体的组成和核心技术；然后探讨了智能体和具身智能的关系，以及实现具身智能所需的关键技术和方法。

第7章：首先介绍AIGC技术的基本概念，然后分别介绍不同类型的AIGC技术的应用场景和案例实践，从中可以让读者体验AIGC的卓越功能及其对我们日常工作和生活的强大助力。

本书适合作为高等院校人工智能、计算机科学及相关专业的教材，也可供从事人工智能技术开发、应用研究的专业人士参考学习。对于广大人工智能爱好者来说，本书同样是一部理想的入门读物。

本书提供电子课件和素材文件，读者可以通过扫描下方的二维码下载相关资料。扫描正文中的二维码可以观看与本书配套的教学视频。

扫码推送配套资源到邮箱

本书由伊犁师范大学的马晓燕撰写并统稿。由于作者水平有限，本书难免有不足之处，欢迎广大读者批评指正。我们的邮箱是992116@qq.com，电话是010-62796045。

作者

2026年1月

目录

第 1 章 人工智能概述 …… 1

1.1 什么是人工智能 …… 1
1.1.1 人工智能的概念 …… 1
1.1.2 人工智能的类型 …… 2
1.1.3 人工智能的特性 …… 5
1.2 人工智能的发展历程 …… 11
1.2.1 图灵测试 …… 11
1.2.2 人工智能的诞生 …… 12
1.2.3 人工智能的3次发展浪潮 …… 12
1.2.4 生成式人工智能的兴起 …… 14
1.3 人工智能的研究内容 …… 16
1.3.1 计算智能 …… 16
1.3.2 感知智能 …… 17
1.3.3 认知智能 …… 18
1.3.4 行为智能 …… 18
1.3.5 群体智能 …… 19
1.3.6 混合智能 …… 20
1.3.7 情感智能 …… 21
1.3.8 类脑智能 …… 24
1.4 人工智能的关键技术 …… 25
1.4.1 机器学习 …… 25
1.4.2 知识图谱 …… 25
1.4.3 人机智能交互 …… 28
1.4.4 自然语言处理 …… 30
1.4.5 计算机视觉 …… 31
1.4.6 生物特征识别 …… 33
1.5 本章小结 …… 35
1.6 习题 …… 36

第 2 章 人工智能的应用 …… 37

2.1 智能制造 …… 37
2.1.1 辅助设计 …… 38
2.1.2 工业仿真 …… 39
2.1.3 智能排产 …… 39
2.1.4 质量管理 …… 40
2.1.5 设备管理 …… 42
2.1.6 智慧物流 …… 43
2.1.7 生产线自动化 …… 44
2.2 医疗健康 …… 45
2.3 金融行业 …… 46
2.4 智慧城市 …… 48
2.4.1 城市生命线维护 …… 49
2.4.2 城市交通管理 …… 49
2.4.3 城市安全治理 …… 50
2.5 智能客服 …… 50
2.6 本章小结 …… 53
2.7 习题 …… 54

第 3 章 机器学习技术 …… 55

3.1 什么是机器学习 …… 55
3.2 机器学习的研究对象 …… 56
3.3 机器学习的主要任务 …… 56
3.3.1 分类 …… 56
3.3.2 回归 …… 57
3.3.3 聚类 …… 58
3.3.4 降维 …… 60
3.4 机器学习的方法 …… 62

3.4.1 有监督学习 …… 62
3.4.2 无监督学习 …… 63
3.4.3 半监督学习 …… 65
3.4.4 强化学习 …… 66
3.4.5 自监督学习 …… 68
3.5 机器学习的工作流程 …… 71
3.5.1 问题定义与数据探索 …… 72
3.5.2 数据预处理 …… 73
3.5.3 模型构建与训练 …… 74
3.5.4 模型评估与结果分析 …… 74
3.5.5 模型部署与应用 …… 75
3.6 机器学习的三大要素 …… 75
3.6.1 模型 …… 76
3.6.2 策略 …… 77
3.6.3 算法 …… 79
3.7 本章小结 …… 84
3.8 习题 …… 84

第 4 章 深度学习技术 …… 87

4.1 什么是深度学习 …… 87
4.2 深度学习与机器学习的区别 …… 88
4.3 深度学习的核心技术 …… 89
4.3.1 Transformer架构 …… 90
4.3.2 预训练大模型 …… 95
4.3.3 预训练大模型的闭源与开源 …… 98
4.4 深度学习的科研应用 …… 101
4.4.1 AlphaFold的革命性突破 …… 102
4.4.2 AlphaFold的技术实现 …… 104
4.4.3 AlphaFold的影响与应用 …… 106
4.5 本章小结 …… 107
4.6 习题 …… 107

第 5 章 大语言模型概述 …… 109

5.1 什么是大语言模型 …… 109
5.1.1 大语言模型的功能 …… 109
5.1.2 大语言模型的特点 …… 110
5.1.3 大语言模型的应用模式 …… 111
5.1.4 人工智能和大语言模型的关系 …… 113
5.2 主流大语言模型 …… 113
5.2.1 国外大语言模型 …… 113
5.2.2 国内大语言模型 …… 115
5.3 大语言模型的智能实现 …… 117
5.3.1 大语言模型的重要概念 …… 117
5.3.2 注意力机制 …… 123
5.4 大语言模型的工程实现 …… 129
5.4.1 规模扩展 …… 129
5.4.2 数据工程 …… 131
5.4.3 大语言模型的能力层级 …… 132
5.5 大语言模型的应用开发技术 …… 133
5.5.1 大语言模型的应用模式 …… 133
5.5.2 大语言模型的微调技术 …… 136
5.5.3 检索增强生成(RAG) …… 139
5.6 大语言模型的评估 …… 144
5.6.1 模型评估面临的挑战 …… 144
5.6.2 模型评估的标准 …… 145
5.6.3 模型评估的分类 …… 146
5.7 本章小结 …… 147
5.8 习题 …… 147

第 6 章 智能体与具身智能 …… 149

6.1 智能体 …… 149
6.1.1 什么是智能体 …… 150
6.1.2 智能体的应用场景与优势 …… 151
6.1.3 智能体的关键特征 …… 152
6.1.4 智能体的分级 …… 153
6.1.5 智能体的组成与工作原理 …… 154
6.1.6 智能体的核心技术与典型产品 …… 155
6.1.7 智能体的发展前景 …… 156
6.2 具身智能 …… 157
6.2.1 什么是具身智能 …… 157
6.2.2 具身智能和智能体的关系 …… 158
6.2.3 具身智能的核心技术 …… 159
6.2.4 具身智能的应用领域 …… 160
6.2.5 人形机器人 …… 162

6.3 本章小结 …… 163
6.4 习题 …… 164

第 7 章 AIGC应用与实践 …… 165

7.1 认识AIGC …… 165
7.1.1 AIGC的概念与特点 …… 165
7.1.2 从人工智能到AIGC …… 166
7.1.3 大模型与AIGC的关系 …… 167
7.2 AIGC内容生成 …… 168
7.2.1 内容生成的种类 …… 168
7.2.2 内容生成的工具 …… 172
7.2.3 行业应用领域 …… 175
7.3 AIGC与文本类写作 …… 178
7.3.1 AIGC的写作流程 …… 178
7.3.2 认识提示词 …… 179
7.3.3 提示词的设计原则 …… 179
7.3.4 提示词优化方法 …… 180
7.3.5 文本类写作案例解析 …… 181
7.4 AIGC与高效办公 …… 185
7.4.1 文档内容创作与编辑 …… 185
7.4.2 表格数据处理 …… 187
7.4.3 智能PPT演示文稿 …… 189
7.5 AIGC与图像处理 …… 191
7.5.1 文本到图像的生成 …… 191
7.5.2 图像到图像的转换 …… 195
7.5.3 AIGC在设计行业中的应用 …… 197
7.6 AIGC与音频编辑 …… 202
7.6.1 AIGC在语音类的应用 …… 202
7.6.2 使用魔音工坊生成语音 …… 203
7.6.3 使用剪映专业版进行AI配音 …… 204
7.7 AIGC与视频创作 …… 207
7.7.1 AIGC在视频类的应用 …… 207
7.7.2 使用剪映专业版生成视频 …… 207
7.8 AIGC与新媒体运营 …… 210
7.8.1 新媒体运营内容的创作与编辑 …… 210
7.8.2 AIGC在新媒体运营中的应用 …… 211
7.8.3 使用AIGC编写媒体宣传文案 …… 212
7.8.4 使用AIGC抓取热点和分析趋势 …… 213
7.9 AIGC与程序设计 …… 214
7.9.1 AIGC在辅助编程中的应用 …… 214
7.9.2 使用AIGC辅助编写代码 …… 215
7.9.3 使用AIGC辅助注释代码 …… 217
7.10 本章小结 …… 220
7.11 习题 …… 220

第 1 章

人工智能概述

本章首先对人工智能的概念进行探讨，明确其技术要素与科学基础，梳理其发展历程并展望技术方向，以帮助读者构建系统性的认识。

本章重点

- 了解人工智能的多维度定义、发展脉络及核心科学基础
- 掌握人工智能的主要研究内容、技术分类与核心特性
- 理解机器学习、计算机视觉等关键技术的基本原理与典型应用

1.1 什么是人工智能

我们自称为“智人”(意为“有智慧的人”)，正是因为“智能”在我们的生存与发展中占据核心地位。数千年来，人类始终在努力理解自身如何思考与行动，即探索我们的大脑如何凭借有限的物质，去感知、理解、预测并操控一个远比自身庞大和复杂的世界。而人工智能(Artificial Intelligence，AI)正是这一探索历程在技术时代的延续与拓展。

1.1.1 人工智能的概念

在1956年的达特茅斯会议上，当时麻省理工学院的约翰·麦卡锡(John McCarthy)首次提出了“人工智能”的概念：人工智能就是要让机器的行为看起来就像是人所表现出的智能行为一样。这一定义不仅为人工智能领域奠定了基础，也引发了广泛的研究和讨论，推动了人工智能领域的发展。

在人工智能的发展历程中，研究者们曾经从不同角度对智能进行了定义和探索，形成了几种具有代表性的观点。其中，一部分研究者将智能定义为对人类行为的模仿，认为机

器若能展现出与人类相似的行为模式，即可被视为具有智能。另一部分研究者则倾向于从更为抽象和正式的角度出发，以“理性”来定义智能，其核心在于做“正确的事情”。这两种定义方式分别对应着“人”与“理性”两个维度。

进一步地，对于智能本身的探讨也呈现出不同的侧重点。一些研究者将智能视为一种内部思维过程和推理的属性，强调智能的本质在于思考和决策的过程。而另一些研究者则更关注智能的外部表现，即智能行为本身。这两种视角分别对应着“思想”与“行为”两个维度。

综合“人与理性”和“思想与行为”这两个维度，可以衍生出4种组合。目前，对人工智能的定义大多可以归入这4种组合之中。用通俗易懂的语言来表述，这4种组合分别是：机器“像人一样行动”、机器“像人一样思考”、机器“理性地思考”及机器“理性地行动”(在这里，“行动”一词应被广义地理解为采取行动或制定行动决策的过程，而并非单纯指肢体动作)。

这4类定义各有其追随者和相应的研究项目，如表1-1所示。

表1-1

类型	研究项目	说明
类人行为	图灵测试方法	这种方法通过测试机器是否能够在对话中欺骗人类评委，使其相信自己是人类，来评估机器的智能水平
类人思考	认知建模方法	这种方法通过建立数学模型来模拟人类的认知机制，从而实现类人思考。例如，通过模拟人类的神经网络结构和信息处理过程，研究者们可以开发出能够进行类似人类推理和学习的算法
理性思考	“思维法则”方法	这种方法认为：智能的本质在于遵循逻辑规则和推理法则，从而做出合理的决策。例如，专家系统就是一种典型的“思维法则”方法，它通过编码人类专家的知识和推理规则，使机器能够解决各领域的复杂问题
理性行为	理性智能体方法	这种方法不强调对人类行为的直接模仿，而是关注机器在复杂环境中实现目标的能力。例如，强化学习算法就是一种理性智能体方法，它通过让机器在环境中试错，学习最优的行为策略

表1-1中所述4种方法各有其独特之处和必然的差异。追求类人智能的方法本质上是一种与心理学紧密相连的经验科学，它依赖于对人类真实行为和思维过程的细致观察与合理假设。相对而言，理性主义方法则是数学与工程学的深度融合，并且与统计学、控制理论以及经济学等领域有着广泛的联系。对这4种方法有了基本的了解之后，便可以构建起对人工智能的初步概念框架。

构建人工智能的初步概念框架

1.1.2 人工智能的类型

按照功能目标分类，我们可以将人工智能划分为“弱人工智能”“强人工智能”“超

人工智能”，如表1-2所示(目前，大多数人工智能系统属于弱人工智能，它们能够在特定领域高效地完成既定任务，但缺乏广泛的自我意识和自主决策能力)。

表1-2

类型	特点	现状	与人类智能对比
弱人工智能	专注于单一任务，无自主意识。典型应用是图像识别(人脸识别)、语音助手(例如Siri)、推荐系统(抖音、淘宝算法)	已广泛应用	低于人类水平
强人工智能	也称为“通用人工智能”，具备人类水平的通用认知能力。例如OpenAI的GPT系列向AGI演进的尝试	研究中	接近人类水平
超人工智能	全方位超越人类智能的形态	理论阶段	远超人类水平

按能力等级分类，人工智能可以分为感知智能、认知智能和创造智能，如表1-3所示。

表1-3

类型	代表技术	典型应用
感知智能	计算机视觉、语音识别	人脸识别、语音转文字
认知智能	知识图谱、推理引擎	法律咨询AI、医疗诊断辅助
创造智能	生成式AI	AI绘画、AI写作

按学习方式分类，人工智能可以分为有监督学习、无监督学习、强化学习、半监督学习、自监督学习，如表1-4所示。

表1-4

类型	原理	典型应用
有监督学习	使用已标注的数据进行训练，学习从输入到输出的映射关系	分类(如垃圾邮件识别)、回归(如房价预测)
无监督学习	分析未标注的数据，自动发现其中的隐藏模式或内在结构	聚类(如客户分群)、降维(如数据可视化)
强化学习	智能体通过与环境交互，根据获得的奖励或惩罚调整行为策略	游戏AI(如AlphaStar)、机器人控制
半监督学习	利用少量标注数据和大量未标注数据进行训练，以降低对标注数据的依赖	医学影像分析、文本分类(标注成本高的任务)
自监督学习	利用数据本身的结构或属性自动生成标签(伪标签)来训练模型，学习数据的有用表示，通常无须人工标注数据	自然语言处理预训练(如BERT、GPT)、计算机视觉预训练(对比学习)、语音表示学习、推荐系统特征学习

按照应用领域来划分，人工智能的应用范围广泛，涵盖了医疗健康、金融投资、智能制造及智慧城市等多个关键领域，如表1-5所示。

表1-5

类型	关键技术	代表产品/系统
医疗健康	医学影像分析、药物分子生成	DeepMind AlphaFold(蛋白质结构预测)
金融投资	欺诈检测、量化交易	蚂蚁集团智能风控系统
智能制造	工业质检、预测性维护	特斯拉工厂AI视觉检测系统
智慧城市	交通流量预测、智能安防监控	海康威视城市大脑平台

按系统架构分类，人工智能可以分为单体AI系统、分布式AI系统、边缘AI(Edge AI)及云端AI，如表1-6所示。

表1-6

类型	特点	典型应用
单体AI系统	由单一模型独立完成特定任务，通常部署于本地环境中	单机版图像分类软件、离线语音识别工具
分布式AI系统	多个模型或模块协同工作，通过系统集成实现复杂任务	自动驾驶系统(感知+决策+控制模块联动)
边缘AI	模型部署在终端设备本地运行，低延迟、高隐私保护，适应实时性要求高的场景	手机人脸解锁、无人机自主避障、工业传感器实时诊断
云端AI	依赖云计算平台提供算力和存储，支持大规模模型训练与推断	DeepSeek、ChatGPT、大模型训练、超算辅助药物研发

按照AI系统的智能形态来分类，人工智能可以分为生成式AI、具身智能及AI Agent(人工智能代理)，如表1-7所示。

表1-7

类型	关键技术/架构	典型应用
生成式AI	扩散模型(如Stable Diffusion)、大语言模型(LLM)	DeepSeek、ChatGPT、MidJourney、DALL-E
具身智能	机器人技术、多模态感知、运动控制	波士顿动力Atlas、宇树机器人
AI Agent(人工智能代理)	感知→规划→行动闭环架构	AutoGPT、Microsoft AutoGen

了解人工智能的类型不仅有助于我们更好地理解人工智能技术的全貌，还能为实际应用、研究方向的选择以及未来的发展规划提供指导。

1.1.3　人工智能的特性

从功能与应用的角度来看，人工智能展现出多方面的核心特性。首先体现为自动化，即AI系统能够在无人持续干预的情况下自主执行任务；其次是适应性，系统能够从经验与数据中学习，并持续优化自身性能；第三是交互性，例如智能客服系统能够以自然方式与人类用户进行交流。此外，人工智能还具有鲁棒性，即使在环境不确定或输入变化的情况下，仍能保持稳定可靠的运行表现。

为了更好地理解人工智能的特性，我们有必要厘清人工智能与其他相关技术和概念之间的关系。接下来，我们将探讨几个常与人工智能紧密相连或容易混淆的概念，包括知识工程、脑机接口、机器学习、深度学习、机器人技术、自动化系统等。

1. 知识工程与人工智能

1977年，美国计算机科学家爱德华·费根鲍姆(Edward Feigenbaum)正式提出并定义了“知识工程”。他在1994年获得了图灵奖，被誉为专家系统之父和知识工程的奠基人。知识工程是一种自上而下的方法，依赖于专家的干预，其基本目标是将专家的知识赋予机器，利用机器来解决问题。

知识工程的起源可追溯至传统的知识管理系统，其发展历程可以划分为图1-1所示的3个阶段。

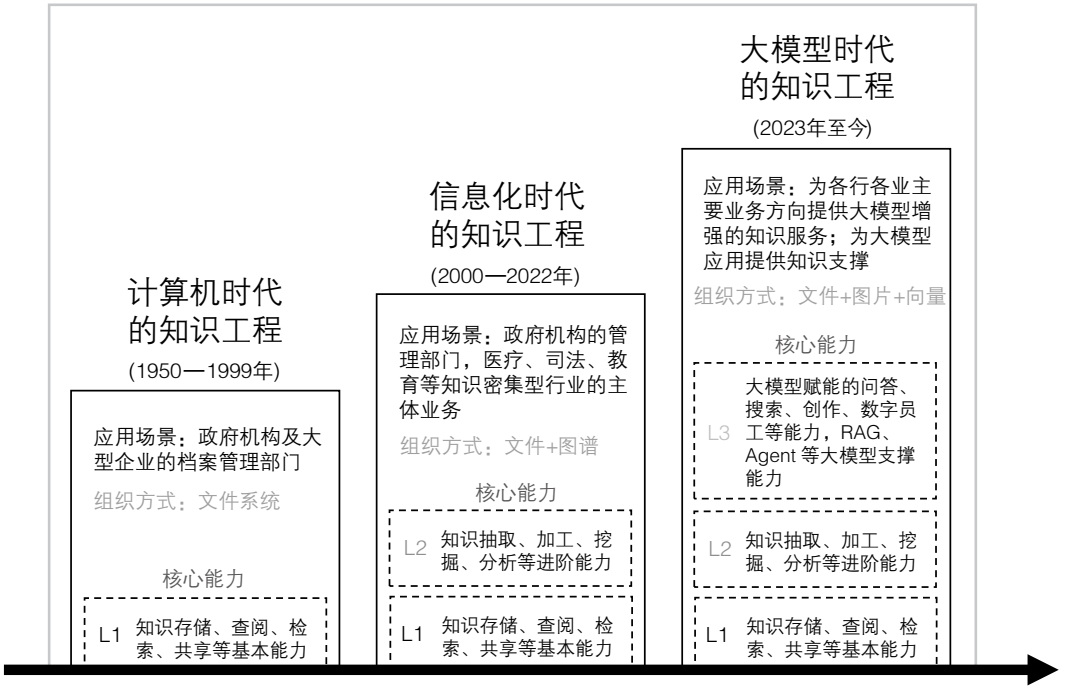

图1-1

20世纪80年代，知识工程的核心应用——专家系统，曾一度风靡。这类系统通过模拟领域专家的推理过程，为各类复杂问题提供决策支持。其典型工作流程包括：首先从专家

处获取知识并构建知识库，随后通过人机接口接收用户的咨询信息，借助推理机对知识库中的规则进行逻辑推理，最终向用户返回建议或答案，如图1-2所示。

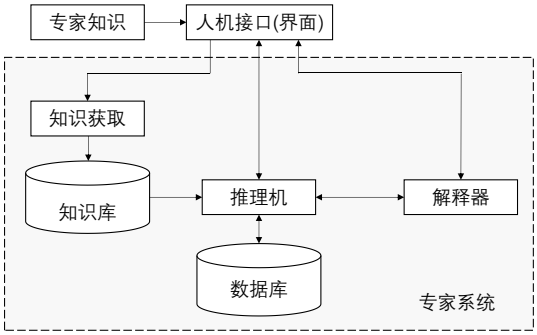

图1-2

同时，专家系统也是早期人工智能的重要分支，可以被视为具备专业知识和经验的计算机智能程序系统。这类系统展现了人工智能的一些基本特性，例如，专家系统能够基于内置的知识库和推理机制，自动完成复杂的决策任务(如医疗诊断和故障分析)，无须人类的持续干预。此外，专家系统通常提供友好的人机交互接口，使用户能够输入问题并获得系统生成的解释和建议(例如，早期的医疗咨询系统便充分体现了这一典型特征)。

在过去的数十年中，知识工程的研究持续推进，专家系统在理论与技术方面不断成熟，其应用已广泛渗透到化学、数学、物理、生物、医学、农业、气象、地质勘探、军事、工程技术、法律、商业、空间技术、自动控制以及计算机设计与制造等众多领域。其间开发了数千个专家系统，其中许多系统在功能上已达到甚至超过相应领域人类专家的水平，并在实际应用中带来了显著的经济效益。

专家系统的应用

2. 脑机接口与人工智能

脑机接口(Brain-Computer Interface，BCI)是一项革命性技术，能够直接将大脑与外部设备连接起来。通过采集和解析神经信号，BCI实现了“脑”与“机”之间的双向信息交互，完全不依赖传统的肌肉或语言系统。脑机接口技术的示意图如图1-3所示。

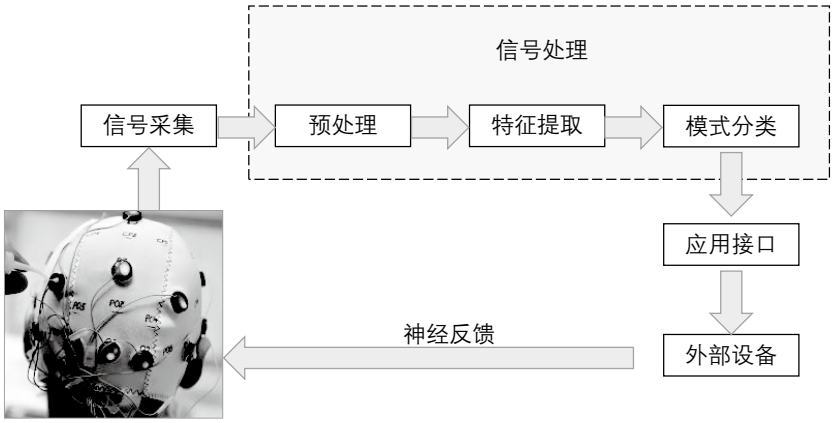

图1-3

在大多数传统脑机接口中，无论是用于控制计算机光标还是机械臂，唯一的控制源均源自解码后的大脑神经电信号。人类执行的诸多动作皆为目标导向型，即动作旨在朝向目标位置展开，例如点击计算机屏幕中的搜索栏、按钮和图标，或者拿起桌上的杯子、钥匙，拉动门把手等。在这些情境下，使用者会在有限数量的可能目标之间权衡，一旦目标确定，其行为往往趋于模式化。此时，人工智能可通过解读意图和辅助行动，显著提升脑机接口使用者的操控力。这种协作模式被称作“共享自主权”，即人类作为主驾，人工智能作为副驾，二者共同操控，从而实现脑机接口性能的大幅提升。

3. 机器学习、深度学习与人工智能

人工智能、机器学习(ML)和深度学习(DL)是3个紧密相关但范围不同的概念，它们之间呈现一种逐层嵌套的关系，如图1-4所示。

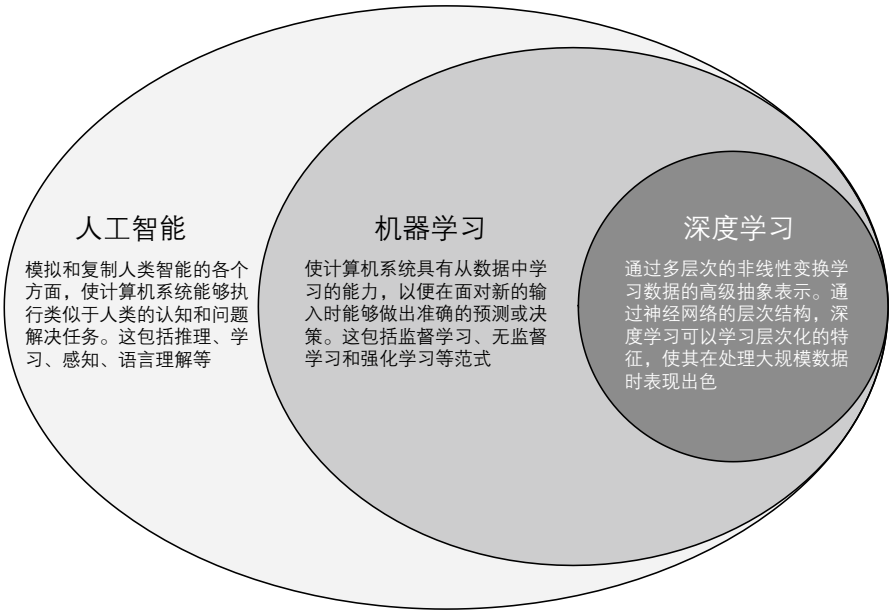

图1-4

从图1-4可以看出，人工智能是一个涵盖范围广泛的概念，泛指一切使计算机系统表现出智能行为的技术，包括规则系统和专家系统等。机器学习作为实现人工智能的一种重要途径，侧重于通过数据驱动的方式不断提升系统性能。而深度学习则是机器学习的一个分支，其核心在于借助深层神经网络结构，实现对数据中高层次抽象特征的自动学习与表示。总体而言，人工智能致力于全面模拟人类的智能行为，机器学习为该目标提供了数据驱动的实现方法，而深度学习则通过在复杂数据中学习多层次表征，推动智能系统性能的提升。

4. 自动化系统与人工智能

人工智能与自动化系统是两个常被提及且容易混淆的概念，尽管它们在表面上有些相似，本质上却存在显著区别。

(1) 自动化系统。自动化系统可被比作一台高度规范的“执行机器”，其严格遵循预设的规则与流程完成指定任务，核心特征体现为高度的重复性和确定性。换言之，自动化系统仅能执行既定的程序指令，不具备自主思考或学习的能力。例如以下情况。

- 汽车制造厂的机械臂：能够以极高精度完成焊接或喷漆作业，但其每一个动作均依据程序员预先设定的指令执行，自身无法做出任何偏离程序的决策。
- 家庭扫地机器人：可按照预设路线进行地面清扫，但它无法主动识别污渍较重区域，也无法通过学习优化清扫策略。

自动化系统的优势在于高效性与准确性，尤其适用于规则明确、重复性高的任务。然而，其局限性在于灵活性的缺失，难以适应复杂多变或未曾预见的场景。

(2) 人工智能。人工智能可以被看作一位“具备学习能力的助手”。它不仅能够执行预定任务，还能借助数据分析不断学习与优化。人工智能的核心特性在于其高度的适应性和智能性，使其能够应对复杂场景，甚至自主做出决策。例如以下情况。

- 语音助手(如Siri或Alexa)：除了回答用户问题，还能根据用户的使用习惯主动推荐音乐或提醒用户日程安排。它们通过持续学习用户行为，逐渐提升个性化服务水平。
- 推荐系统(如Netflix或淘宝)：基于用户的观看记录或购买历史，动态生成符合其偏好的电影或商品推荐。这些推荐并非固定不变，而是通过分析海量数据实时优化得出。

人工智能的优势体现在处理不确定性、适应环境变化乃至生成创新解决方案等方面。然而，其开发与运行通常依赖大规模数据及强劲的计算资源，因此也带来较高的研发与维护成本。

综上所述，自动化系统和人工智能的特性的区别如表1-8所示。

表1-8

特性	自动化系统	人工智能
核心能力	执行预设任务	学习、推理、决策
灵活性	低(只能按规则执行)	高(可以适应新情况)
学习能力	无	有
使用场景	重复性高、规则明确的任务	复杂、不确定性高的任务
案例	工厂机械臂、普通扫地机器人	语音助手、推荐系统

在实际应用场景中，尽管自动化与人工智能在本质上有所不同，但它们并非相互对立，而是能够有效互补，共同提升系统的整体效能。例如以下情况。

- 智能工厂：自动化设备可靠地执行流水线上的重复性任务，而AI系统则实时监控设备状态、预测潜在故障，并基于数据分析优化生产流程。
- 自动驾驶汽车：车辆的加速、刹车等基础操作由高精度的自动化系统负责，而AI则承担感知任务，如识别道路中的行人、车辆及其他障碍物，并据此做出综合驾驶决策。

5. 智能体与人工智能

智能体，也称为“人工智能代理(AI Agent)”，是一种能够在特定环境中自主感知信息、做出决策并执行动作的人工智能系统。它代表了人工智能领域的重要进展，其核心在于能够构建出可自主与环境交互、通过决策与行动实现预设目标的智能实体。

智能体能够在许多场景中发挥关键作用，例如：

- 在客户服务场景中，智能体能够通过分析用户行为和偏好，提供定制化的产品推荐，如电商网站和教育平台的个性化服务。
- 在财务报告与审计场景中，智能体可以自动执行数据输入、文档审查和对账等任务，审计人员将获得更准确的数据，同时从数据验证的烦琐工作中解放出来，将精力转向风险评估和战略规划等更有价值的活动。
- 在自动化安全测试场景中，智能体可以模拟各种攻击场景，对系统进行全面的安全评估。智能体通过分析项目文档和接口规范，生成具有上下文感知能力的测试用例，确保覆盖系统的关键部分，从而显著节省开发时间。
- 在自动化文档处理场景中，智能体能够读取、解析和理解不同格式的文档，识别其中的关键信息并进行自动分类和处理。例如，在法律领域的文档处理中，智能体可以自动分析合同文档，提取合同条款、签约方信息、有效期等关键信息，帮助法律专业人员更快地审查和归档合同。

人工智能为智能体的行为提供了核心技术支撑。通过集成人工智能技术，智能体显著增强了其在决策制定、持续学习与环境适应等方面的能力。从本质上讲，智能体可被视为人工智能的一种具体应用形态，它依赖人工智能的模型与算法来执行复杂任务、积累经验并进行有效的逻辑推理。

6. 具身智能与人工智能

具身智能(Embodied Intelligence)是人工智能的一个前沿分支，强调智能体通过物理身体与真实环境进行交互，并在交互中学习、推理和进化。其实质是人工智能与机器人技术的深度融合：人工智能为机器人提供“大脑”，使其具备感知、思考和决策的能力；而机器人则为人工智能提供“身体”，使其能够与真实世界互动，获取经验和知识。

在应用领域，具身智能凭借其独特的技术优势，已在众多场景中实现广泛应用，如图1-5所示，其应用范围不仅限于机器人，还覆盖了eVTOL(电动垂直起降飞行器)、无人机、智能驾驶汽车等多个重要方向。

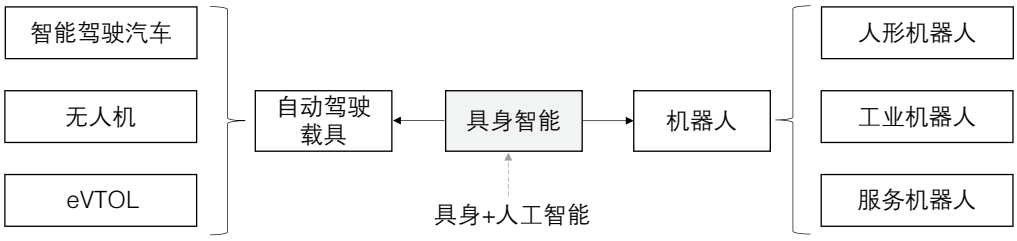

图1-5

具身智能与前面介绍过的智能体是人工智能领域中两个密切相关但有所区别的概念，其核心差异体现在对物理“身体”的依赖程度及与环境交互的方式(如表1-9所示)。总体而言，具身智能可视为智能体的一个子集：所有具身智能体都属于智能体，但并非所有智能体都具备“具身性”。具身性代表一种更强的约束条件，它在传统智能体的基础上，进一步要求智能体必须拥有物理实体，并能够与真实环境进行实时交互。

表1-9

维度	智能体	具身智能
身体依赖	不要求物理身体(可以是虚拟的)	必须拥有物理身体(如机器人、智能汽车、无人机)
环境交互方式	可仅通过数据或符号与环境交互(如输入文本、图像)	必须通过物理动作(如移动、触摸、力学反馈)与环境互动
智能来源	可能依赖纯算法或数据驱动(如深度学习)	身体感知与行动闭环是智能的核心(如通过跌倒来学会平衡)
适应性	适应虚拟或结构化环境(如游戏AI)	适应复杂物理环境(如动态地形、温度变化)

7. Python与人工智能

人工智能程序可以使用多种编程语言(如 C/C++、Java 等)实现，但 Python 却在人工智能领域被广泛视为首选语言。那么，究竟是什么原因使得 Python 在众多选项中脱颖而出呢？

首先，Python 作为美国众多大学中广受欢迎的入门编程语言，诞生至今已有三十余年。与其他编程语言相比，Python 更加易学、易读，非常适合快速开发。其编程风格简单、直接，语言难度低于 Java 等语言，对初学者尤为友好，能够帮助开发者更专注于解决问题，而不必过多纠结于语法细节。

其次，在人工智能领域中，数据如同工厂的“原材料”，其质量直接决定最终产物，即模型性能的优劣。原始数据往往包含诸多问题，例如存在缺失值、异常数据点或格式不一致等情况。此时，Python生态中的Pandas和NumPy等库可以扮演“数据整理大师”的角色。Pandas能够高效读取多种格式的数据文件(如CSV和Excel文件)，并轻松应对缺失值与异常值的处理，同时支持灵活的数据格式转换。NumPy 则提供了强大的数值计算功能，可高效执行各类数学运算与统计分析。借助这些工具的处理，数据得以变得整洁、规范，为后续人工智能模型的训练奠定了可靠的基础。

再次，Python的机器学习和深度学习库是构建人工智能模型的核心工具。

(1) 在机器学习领域，Scikit-learn作为常用的机器学习库，集成了众多经典算法，如线性回归、决策树和支持向量机等。即便是初学者，也能够通过简单操作调用这些算法完成模型的构建与训练，使模型具备基础的“思考”和“预测”能力。

(2) 在深度学习领域，TensorFlow和PyTorch作为两大主流框架发挥着关键作用。TensorFlow由谷歌开发，具备强大的计算性能和分布式训练支持，尤其适用于处理大规模数据的深度学习任务；PyTorch则凭借其动态计算图和易于调试的特性，广受开发者青睐。借助这些框架，用户能够构建复杂的神经网络模型，为人工智能系统注入更强大的“智慧”。

Python编程入门

1.2　人工智能的发展历程

人工智能技术自诞生以来，其发展历程经历了多次起伏，从早期的符号主义推理到当前蓬勃发展的生成式AI，每一次技术突破都深刻改变了人类对智能的理解与应用边界。

1.2.1　图灵测试

1950年，“计算机之父”与“人工智能之父”艾伦·图灵(Alan Turing)发表了论文《计算机器与智能》，该文被公认为人工智能领域的开山之作。在论文开头，图灵提出了一个至今仍引人深思的问题：“机器能思考吗？”这一问题不仅激发了人们对智能机器的无限想象，也奠定了人工智能研究的基本概念与理论雏形。

在这篇论文中，图灵提出了一种判断机器是否具备智能的方法，即后来成为人工智能领域经典标准的“图灵测试”。如图1-6所示，该测试的基本流程为：测试者在与被测试者(一个人和一台机器)完全隔离的情况下，通过键盘等设备随意向双方提问。经过多轮测试后，如果机器能够令测试者平均误判率超过30%，即被认为通过测试，并判定该机器具有人类水平的智能。

图1-6

1.2.2 人工智能的诞生

人工智能的诞生可追溯至20世纪50年代。当时，计算机科学尚处于起步阶段，研究者们已开始尝试通过编写程序来模拟人类的思维与行为。在这一背景下，一批杰出的科学家与工程师积极投身于如何赋予计算机更高级智能功能的探索之中。

1956年夏季，人工智能领域迎来了一场具有里程碑意义的会议——达特茅斯会议。此次历史性盛会汇聚了当时顶尖的计算机科学家与认知学者，与会者围绕“构建能够模拟人类思考与学习的机器”这一共同愿景，进行了为期8周的深入研讨，为人工智能作为一个独立学科的形成与发展奠定了坚实基础。

作为人工智能发展史上的重要转折点，达特茅斯会议为研究者们提供了系统探讨人工智能核心议题与前沿思想的平台。会议集中讨论了图1-7所示的5个主要方向。

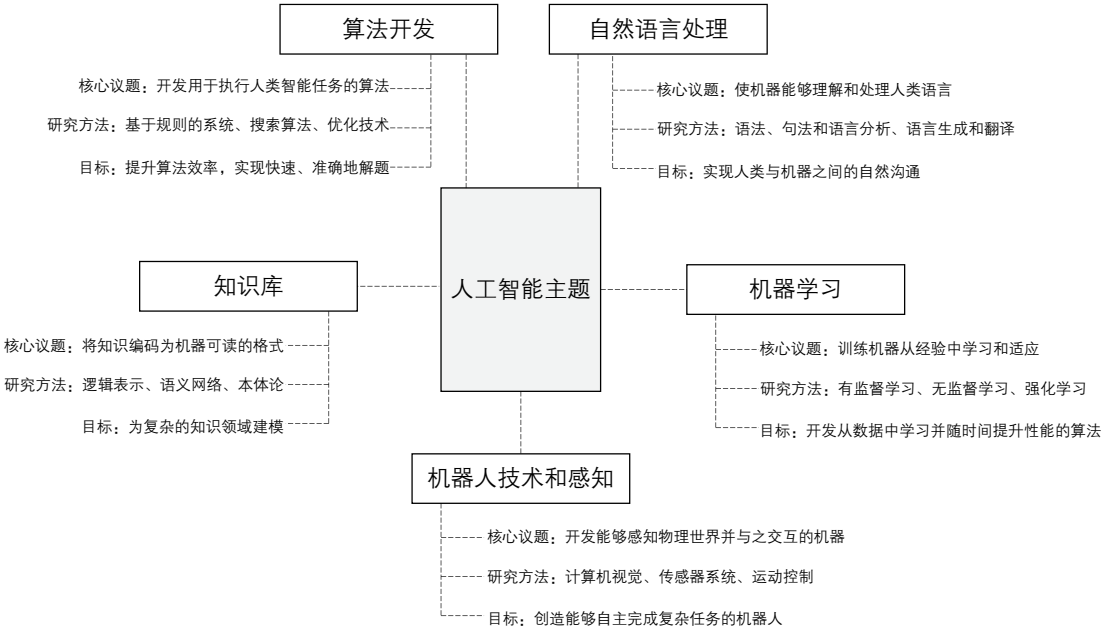

图1-7

在会议上，约翰·麦卡锡(John McCarthy)首次提出“人工智能”一词。达特茅斯会议标志着人工智能作为独立学科的诞生，被称为“人工智能的开端”，1956年也因此被称为“人工智能元年”。

1.2.3 人工智能的3次发展浪潮

达特茅斯会议之后，人工智能领域进入了一个蓬勃发展的新阶段。参与研究的学者规模显著扩大，逐渐形成了多个具有不同理论取向与技术路径的学术派系(如符号主义学派、连接主义学派和行为主义学派)，同时也经历了3次主要的发展浪潮，如图1-8所示。

人工智能的三大学派

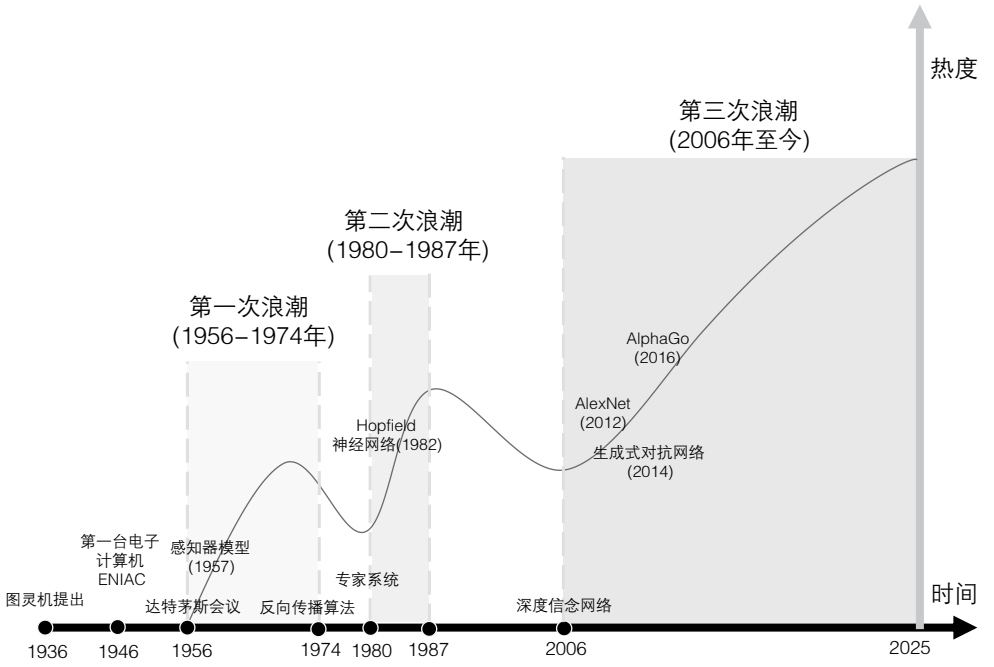

图1-8

1. 第一次浪潮(1956–1974年)

自1956年起，人工智能迎来了第一波发展浪潮，这一阶段一直持续到1974年。当时的研究主流认为机器的智能源于逻辑推理能力，由此提出了符号主义(Symbolism)，将人类认知过程视为符号操作过程。研究者开发了多种启发式算法(Heuristic Algorithm)，如遗传算法、粒子群算法、蚁群算法、禁忌搜索算法和模拟退火算法等，并以此为基础推动了专家系统与知识工程理论的建立。

1957年，弗兰克·罗森布拉特(Frank Rosenblatt)在M-P模型的基础上提出了感知器(Perceptron)模型，该模型能够完成一些简单的视觉处理任务，为后来的神经网络与支持向量机(Support Vector Machine，SVM)奠定了基础。然而在1969年，马文·明斯基(M. L. Minsky)指出感知器本质上是一种线性模型，无法处理非线性问题(如经典的“异或”问题/XOR问题)，这一结论使神经网络研究陷入近二十年的低谷。

随后，由于计算机性能限制、计算复杂性急剧上升以及训练数据匮乏，许多人工智能项目难以推进，理论与算法的发展也遭遇瓶颈，人工智能领域的发展逐渐进入一段低潮期。

2. 第二次浪潮(1980–1987年)

人工智能的第二次浪潮大约始于1980年，持续至1987年左右。在这一阶段，研究者逐渐摆脱了符号主义范式的限制，转而广泛采用基于统计学的方法来应对人工智能中的各类问题。与此同时，连接主义路径逐渐兴起，研究者通过仿生学途径模拟生物神经元的功能，构建并发展了神经网络模型，为该领域注入了新的理论动力与技术方向。

在第二次浪潮中，“连接主义”正式登上历史舞台，从模型构建到算法设计，从理论分析到工程实现，全方位地为人工神经网络(Artificial Neural Network，ANN)计算机走向市场奠定了坚实的基础。从此，机器学习开始自立门户，人工智能也迈入了以机器学习为核心的阶段。在这一时期，人工智能取得了诸多令人瞩目的成就：它接连战胜双陆棋世界冠军、国际象棋世界冠军，知识工程和语音识别领域也实现了重大突破。然而，20世纪 90 年代后，第五代计算机(人工智能计算机)项目以失败告终，同时人工智能在明确的商业应用场景方面也显得捉襟见肘，这些因素共同导致人工智能研究再次陷入困境，进入了第二个低谷期。

3. 第三次浪潮(2006年至今)

人工智能的第三次浪潮始于2006年，至今仍在持续。在这一阶段，连接主义大放异彩，深度学习迅速崛起。20世纪90年代兴起的互联网产业积累了海量数据，为人工智能的复兴奠定了数据基础；与此同时，遵循摩尔定律不断进步的芯片产业为人工智能所需的高性能计算提供了硬件支持。

人工智能产业复兴的号角，是由“神经网络之父”杰弗里·辛顿(Geoffrey Hinton)吹响的。2006年，辛顿发表了具有里程碑意义的论文《一种深度信念网络的快速学习算法》。他提出了一种创新的训练方法：先通过无监督学习方法逐层训练算法(Pre-training)，为网络构建一个良好的初始状态；再使用有监督的BP(反向传播)算法进行调优(Fine Tuning)，对网络进行精细调整。这一方法巧妙地解决了长期困扰深度学习的“梯度消失”问题，为深度学习奠定了坚实的理论基础，也让2006年成为了“深度学习元年”。

2014年，伊恩·古德费罗(Ian Goodfellow)在卷积神经网络(CNN)的基础上提出了“生成对抗网络”(Generative Adversarial Networks，GAN)。这一深度学习模型通过生成器(generator)和判别器(discriminator)的对抗训练，使得两个神经网络互相竞争，前者负责“造假”、后者负责识别，在不断对抗中发展出强大的生成能力，最终得到以假乱真的图片或文字。GAN逐渐成为深度学习模型中最强大的算法之一。

在第三次人工智能浪潮中，融合了连接主义与大数据技术的深度学习算法席卷了整个人工智能领域。借助海量数据训练而成的新一代人工智能系统，不断突破人类对机器能力的想象边界。

1.2.4 生成式人工智能的兴起

随着人工智能技术的持续发展，生成式人工智能(Generative Artificial Intelligence)正逐渐从前沿技术探索走向大规模实际应用，现已成为人工智能领域近年来最具颠覆性与影响力的研究方向之一。

生成式人工智能的核心在于基于海量数据学习内容规律，学会底层的统计规律和特征表示，进而生成新的文本、图像、音频或视频等内容，具有“创作”能力。与强调分类判断的判别式人工智能不同，生成式人工智能建模的是数据的分布特征，并据此生成与原数据相似但不重复的新样本，展现出在写作、绘画、作曲等领域的强大想象力。此外，生成

式人工智能的另一个显著特点是对上下文和风格的敏感性。许多生成式人工智能模型可以接受提示(prompt)或条件输入，按需生成不同风格的内容，如模仿某位作家的文风或绘出指定场景的图像。这使生成式人工智能成为人工智能生成内容(AIGC)的技术基础，在艺术设计、内容创作等方面展现出强大的应用潜力。

表1-10所示为生成式人工智能的常见应用。

表1-10

应用	说明
文本	内容创作：以GPT为代表的生成模型能够自动撰写文章、博客、故事、剧本甚至代码，可显著提升创意与文本生产效率
	对话智能体：基于生成式AI的聊天机器人和虚拟助手能够进行高度拟人化的对话，精准回应用户查询，并提供个性化的信息与服务支持
图像和视频	艺术作品生成与设计：生成对抗网络(GAN)能够生成高度逼真的图像、绘画及其他艺术作品，并在时尚设计、建筑设计等领域提供创意支持与实践辅助
	视频合成：该技术能够合成高质量视频内容，应用范围涵盖从深度伪造模拟到电影与电子游戏中逼真场景的生成
音乐和声音	人工智能模型能够创作原创音乐、生成逼真音效，并设计出富有创意的乐器编曲
合成数据	生成合成数据集可用于有效扩充训练样本，这一应用在医疗保健、金融和自动驾驶等领域中，对提升机器学习模型的训练效果具有重要价值

生成式人工智能的迅猛发展离不开多项关键技术的突破。始于2017年提出的革命性Transformer架构凭借自注意力机制彻底改变了自然语言处理(NLP)领域的发展路径。2018年，首批基于Transformer的大语言模型——GPT和BERT相继发布，显著提升了模型对语境的理解和文本生成能力，为后续技术突破奠定了坚实基础。

2020年，参数量达到1750亿的GPT-3问世，展现出卓越的少样本乃至零样本学习能力。然而，该类模型也暴露出“幻觉”问题，即生成与事实不符的内容，成为实际应用中的关键挑战。为应对这一问题，OpenAI于2022年引入“监督微调”(SFT)和“基于人类反馈的强化学习”(RLHF)等技术，在此基础上开发出对话型模型ChatGPT。这一突破迅速引发全球范围内对人工智能的新一轮广泛关注。

2023年及2024年，多模态模型(如GPT-4和GPT-4o)持续演进，实现了文本、图像与音频处理的深度融合，使模型具备更接近人类的“听”“说”“看”的能力。与此同时，OpenAI的o1和DeepSeek的R1等模型在复杂推理方面取得重要进展，推动大语言模型趋近于类似人类的“系统二思维”。此外，2025年发布的DeepSeek-R1以出色的成本效益和开源策略，对现有AI开发规范提出挑战，大大促进了先进大语言模型的普及应用，为各行各业的创新应用注入了强劲动力。

图1-9所示为生成式人工智能的发展历程。

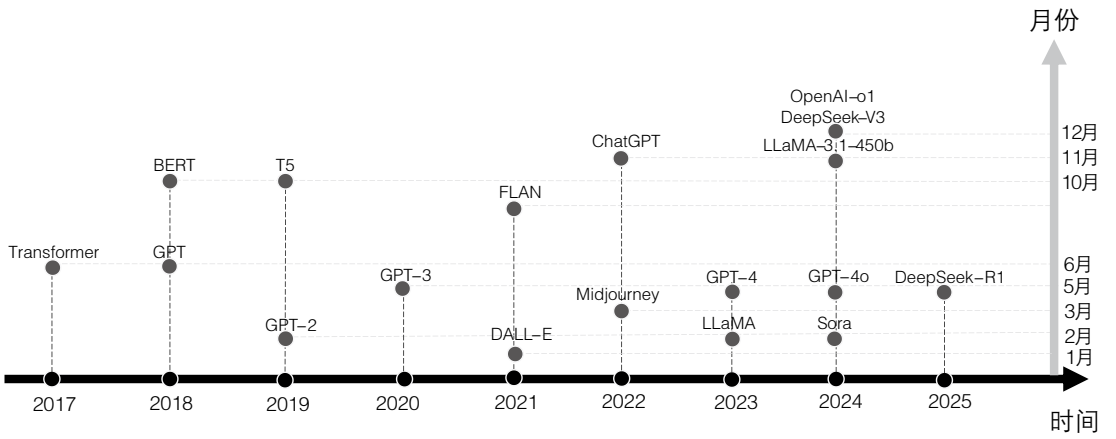

图1-9

当前，生成式人工智能的典型应用已涵盖文本生成(如ChatGPT、DeepSeek)、图像生成(如Midjourney、DALL-E)以及视频生成(如Sora)等多个领域，共同构成了多模态内容生产的核心路径，为各行各业带来了强大的赋能与变革。

人工智能在国内的发展

1.3 人工智能的研究内容

从模拟人类智能的视角来看，人工智能的研究内容可大致划分为以下几个层面：计算智能、感知智能、认知智能、行为智能、群体智能、混合智能、情感智能及类脑智能等。

1.3.1 计算智能

计算智能是人工智能领域中一个重要的前沿研究方向，其核心在于使机器具备高效的计算能力和海量数据存储能力。在这一层面，人工智能系统能够以远超人类的速度和精度，对大规模数据进行处理、分析及复杂数学运算，从而为后续的感知智能与认知智能的发展提供必需的数据支撑和算力资源。

在实际应用中，以下两类技术具有代表性。

- 分布式计算系统：通过多台计算机协同工作，实现大规模数据的并行处理，如云计算平台就是典型的分布式计算系统应用。
- 神经网络模型：模拟人脑神经元之间的连接和信号传递方式，构建复杂的数学模型来处理非线性问题，如深度学习中的卷积神经网络和循环神经网络(RNN)。

计算智能的重要意义在于显著提升了数据处理与计算的效率，为人工智能技术在多个领域的深入应用奠定了坚实基础。它不仅推动了人工智能在金融、医疗和科学研究等重要

领域发挥关键作用，也促进了社会经济的快速发展。

1.3.2　感知智能

感知智能是人工智能发展过程中的第二个关键阶段，它使机器能够具备类似人类的感官认知能力。在这一阶段，人工智能系统可以识别并理解来自外部环境的图像、声音、文字等多种信息，从而实现与外部环境的有效交互。这种感知能力的突破，使得人工智能能够在自动驾驶、具身智能等复杂场景中发挥更加广泛和深入的作用。

- 在自动驾驶领域，智能感知系统通过集成传感器、移动互联网与人工智能技术，实现了对车辆的自主控制，从而显著减少甚至替代传统人工驾驶，主动适应并高效响应用户的出行需求。
- 在具身智能领域，智能感知是智能体与环境交互的第一步，它强调的是主动式、全方位的信息获取与处理。要求智能体搭载多种传感器，并进行多源异构信息的融合与处理，以实现对动态环境的全面理解，如图1-10所示。

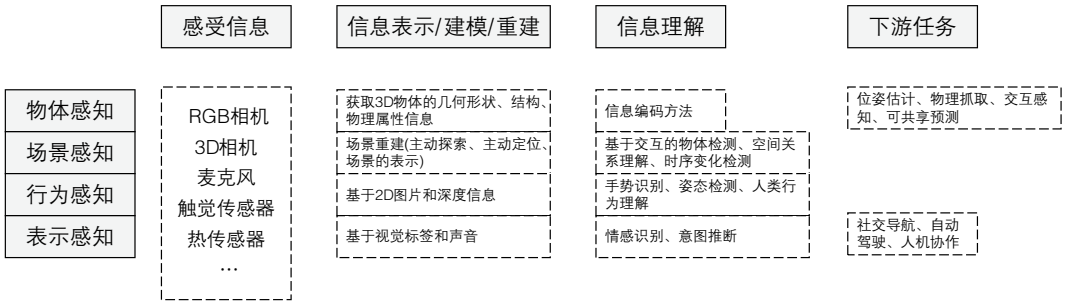

图1-10

感知智能涵盖计算机视觉、语音识别、触觉传感、多模态感知等一系列关键技术。这些技术虽在应用场景与实现方式上各有不同，但共同致力于推动机器向更智能的方向发展，使其能够更准确地理解并响应人类的需求，如表1-11所示。

表1-11

关键技术	核心功能	应用领域
计算机视觉	使机器识别和理解图像/视频，完成检测、分类、分割等任务	自动驾驶、安防监控、工业检测、生物识别(人脸、指纹等)
语音识别	将声音信号转换为文本，处理和理解音频输入	智能助手、客服系统、智能家居
触觉传感	模拟人类触觉，感知物体形状、纹理、硬度等物理属性	机器人手术、精密操作、人机交互、自动驾驶车辆
多模态感知	融合视觉、声音、触觉等多种信号，实现全面环境感知与决策	自动驾驶、智能机器人、增强现实(AR)、人机交互系统

1.3.3 认知智能

认知智能被视为人工智能发展的高级阶段，也是目前尚未完全实现的远景目标。在认知智能领域，未来的发展趋势是训练人工智能，从而使其获取智慧本质的“通式”。具体而言，就是借鉴人类智慧的特点，通过构建多模态大模型，实现知识的积累和智慧的应用。例如：

- 在制造业和物流行业中，存在大量需要分拣的作业环节。若采用人工分拣，不仅作业速度慢、成本高，还需提供适宜的环境以保障人员操作。而采用人工智能工业机器人进行智能分拣，则能够显著提高分拣速度，大幅降低运营成本。
- 认知智能与医疗场景的深度融合，不仅能够基于医疗数据推理诊断逻辑，辅助医生进行疾病判断，还能为患者量身定制个性化治疗方案，从而有效提升治疗效果与患者生存率(认知智能工具可广泛应用于心血管疾病、癌症、糖尿病等多种疾病的诊断与治疗过程中)。
- 认知智能工具能够生成高度仿真的情境，构建科学化的数据环境，使学生通过实验性探索和即时反馈深入理解事物之间的内在联系，从而掌握系统性解决问题的办法。同时，该工具可实时采集学习过程数据，构建动态学习者画像，并针对学生的认知薄弱环节提供多样化的思维支持，例如问题引导、概念强化和案例分析等策略。

认知智能致力于提升机器在理解、思考与推理等方面的能力，其核心目标是使机器能够准确解读语言、把握概念，并完成逻辑推断。随着数据红利逐渐消退和深度学习技术的日益成熟，认知智能正成为人工智能领域的新兴焦点。以自然语言处理和知识图谱为代表的认知智能技术，正在推动机器智能向更高层次发展。

1.3.4 行为智能

行为智能重新定义了人工智能在行为能力方面的内涵与边界。它不仅超越了基础的感知与认知层面，更将智能体(无论是机器还是软件)的推理与决策能力提升到了新的高度。正如人类孩子在成长过程中，不仅需要学会观察世界、理解知识，更要逐步掌握如何行动与决策；行为智能的研究目标，正是为了让智能体能够在复杂多变的环境中，像人类一样灵活应对不确定性，并做出智能且合理的决策。下面是一些典型的行为智能应用案例。

- 宇树机器人、波士顿动力的Atlas等人形机器人，能够完成跑酷、后空翻、搬运货物等复杂任务。机器人的每个动作都是动态平衡、环境适应与力量控制的结果。它需要在摔倒的瞬间做出恢复平衡的反应，这是高级行为智能的极致表现，如图1-11所示。
- 智能客服与对话机器人能够代表用户执行任务，例如要求它“帮我把上周开会提到的总结报告发邮件给张总”。其“行为”智能是指操作软件完成实际工作。

在无人机编队表演与自主导航中，成千上万的无人机在空中自主形成复杂图案，或无人机在森林中自主避障巡线。每架无人机都需要根据邻居的位置和全局目标，实时决策自己的飞行路径和速度(行为)，以避免碰撞并保持队形(这是群体行为智能的典型例子)，如图1-12所示。

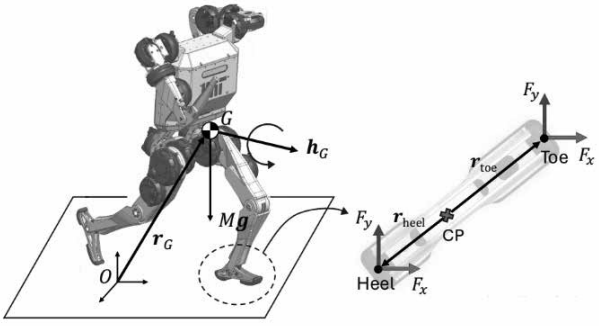

图1-11

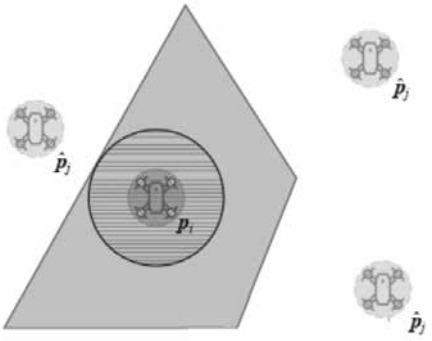

图1-12

1.3.5　群体智能

随着互联网的普及、传感网的渗透、大数据的涌现以及信息社区的崛起，新的信息环境逐渐形成。在这一背景下，数据与信息在人类社会、物理空间和信息空间之间深度融合、广泛交互，深刻改变了人工智能发展所依赖的外部环境与数据基础。人工智能的科学范式和实现路径正面临重大突破，并逐步迈向一个新的发展阶段。这一既源于传统人工智能又显著区别于前者的新范式，被称为“人工智能2.0”(AI 2.0)。

中国工程院于2015年12月批准实施“中国人工智能2.0发展战略研究”重大咨询项目，在国内率先提出并启动“人工智能2.0计划”。项目组完成了《新一代人工智能规划建议研究报告》与《新一代人工智能重大科技项目实施方案》的编制工作，并于2016年明确了人工智能2.0的核心理念。2017年7月，国务院发布《新一代人工智能发展规划》，标志着新一代人工智能的研究与推进已正式上升为国家战略。该规划在基础理论体系中确立了5个主攻方向：大数据智能、跨媒体感知计算、人机混合智能、群体智能、自主协同与决策。

群体智能(Collective Intelligence，CI)，也称为集体智能，是一种区别于个体智能、由群体协同所呈现的智能形态。它源于对自然界中群居生物(如蚁群、鸟群、蜂群等)行为的观察与研究。在该形态下，群体通过个体间的互动与协作，能够涌现出超越任一单个成员能力的更高层次的智能行为，如表1-12所示。

表1-12

特点	说明
分布式控制	无中心控制节点，适应环境变化，鲁棒性强，个体故障不影响整体
共享平台	为完成某目标而行动的这群智能个体皆活动在同一平台上，如某种领域或空间

(续表)

特点	说明
间接通信与可扩充性	通过改变环境实现间接通信，通信开销随个体增加而增长，幅度小
个体简单性	个体能力简单，系统易于设计与实现，低成本，简单性特征显著
自组织与智能涌现	个体局部交互产生复杂全局智能行为，交互产生的涌现效应效果显著
含有共识	智能群体在共同知识库中存储共识。共识会演变，共识的水平是群体水平的体现
自动演化	智能群体会随时间而自动演化。这种演化由群体中智能个体的演化和群体的结构关系演化综合而成

未来的群体智能将走向人与人、人与机器、机器与机器之间相互交织的网络化智能，通过网络将个体智能汇聚在一起，形成远超个体智能的新型智能形态或者超级智能形态。

1.3.6 混合智能

混合智能是人类智能与人工智能的深度融合，旨在充分发挥人类的独特优势，同时通过机器智能弥补人类能力的局限性。如图1-13所示，在人工智能系统中引入受人类认知机制启发的计算技术，实现了人类智能与机器智能的协同结合。该体系以人机智能深度协同为目标，通过建立互联通道，构建了一种能够整合机器与人类在感知、记忆、推理、学习、搜索、计算、优化和存储等多方面能力的新型智能系统。

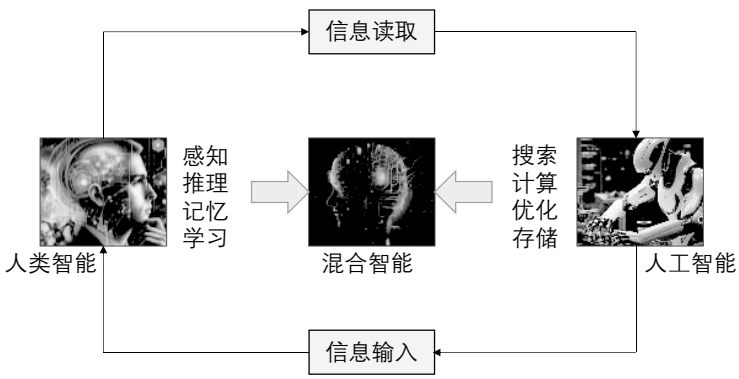

图1-13

人工智能在特定领域的巨大成功，为新一代人工智能的研究与发展提供了重要借鉴和创新路径。在“通用人工智能”真正实现之前，混合智能将作为传统人工智能与未来通用人工智能之间的关键过渡形态。一方面，混合智能借助多样化的人机融合技术，能够增强健全人群、残障人士乃至动物的体能和智能水平；另一方面，通过持续整合人工智能、生物医学工程等跨领域技术，人类有望不断推动人工智能与生物智能的深度融合，从而发展出更为先进的混合智能形态。

表1-13所示为混合智能的应用。

表1-13

应用	说明	典型场景
脑机接口	在大脑与外部设备之间建立直接通信通路，实现意念控制与信息反馈	帮助瘫痪患者控制机械臂、轮椅；通过神经反馈治疗注意力缺陷；未来或可实现直接的大脑交流
可穿戴技术	集成多种传感器的便携设备，可连续、无创地监测人体生理和行为数据	智能手表监测健康；智能眼镜提供实时信息叠加；智能衣物分析运动员姿态和生理状态
可植入电子芯片	植入人体内部的微型电子设备，用于修复、增强或监测特定的生理功能	人工耳蜗恢复听力；人工视网膜让部分失明者产生视觉；神经刺激芯片治疗帕金森病；未来用于记忆增强
外骨骼混合智能	穿戴在人体外部的机械装置，通过AI算法感知人的运动意图，提供助力和保护	减轻搬运工人的负担、防止工伤；帮助中风或脊髓损伤患者进行康复训练；军事上，增强士兵负重和行军能力
人体增强装置	一个更广泛的概念，泛指一切利用技术来超越人类固有生理极限的装置	包括基因编辑、药理学增强等生物技术领域，旨在增强认知、体能或延长寿命

1.3.7　情感智能

情感智能是当前人工智能领域的重要前沿方向，其核心目标是使机器能够识别、理解、响应甚至表达人类情感，从而实现更自然、更富有共情能力的人机交互。目前，情感智能的研究主要涵盖两大领域：情感计算与感性工学。

1. 情感计算

情感信息在人际交流中扮演着关键角色。尽管传统人工智能系统能够高效完成逻辑性强的感知与认知任务，却普遍缺乏对人类情感复杂性的理解能力——它们既无法辨识个体精神异常时的语调起伏，也难以体察人在焦虑状态下的心理需求，更无法解读复杂情境中群体精神状态的多样呈现。这类“情感鸿沟”现象，已成为制约人工智能在服务业等领域深入应用的重要瓶颈。

情感计算技术的兴起，正是为了弥合这一“情感鸿沟”。1997年，美国麻省理工学院罗萨琳德·皮卡德(Rosalind Picard)教授在其开创性著作《情感计算》中首次为该领域做出明确定义，指出情感计算“与情感相关，源自情感或能够对情感施加影响”，从而奠定了这一学科的理论基础。历经近三十年的快速发展与持续迭代，情感计算已成长为人工智能领域中一个特色鲜明的交叉学科：心理学为其提供情感建模的理论支撑；神经科学帮助解

析脑电、皮肤电等生理信号中所蕴含的情感编码机制；计算机科学则致力于通过算法实现对多模态情感信息的精准量化与分析；而哲学与伦理学更在其中扮演价值锚点的角色，为情感计算的有序发展确立伦理框架，既划定技术应用的边界，也确保其演进始终契合人性的本质需求。

如今，情感计算技术已经能够有效地分析多种模态的信号。一方面，它通过面部识别捕捉微表情，通过语音分析提取语速和音调中的情感线索；另一方面，借助可穿戴设备，监测与个体情绪波动密切相关的心率和脑电波等生理指标，甚至能够从文本评论中识别出潜在的情感倾向和主观情绪色彩。在情感计算技术的支持下，新一代人工智能系统正在从“能听会说”向“察言观色”迈进。

例如，通过融合情感计算与3D面部重建技术，人工智能系统可有效排除光照、面部配饰等干扰因素，研究发现，双相障碍患者与抑郁症患者在面部细微运动上存在可量化的关键差异区域：双相障碍患者在外眼角、眉轮廓及外唇轮廓区域表现出显著更高的激活值；而抑郁症患者的识别核心特征则在于内眼角运动的变化。相比之下，健康被试的面部16个区域的激活值分布更为均匀，无明显高激活值的聚焦区域，体现了中性、自然的表情特征。这些肉眼难以察觉的细微面部运动差异，可被AI技术精准捕捉和量化，如图1-14所示。

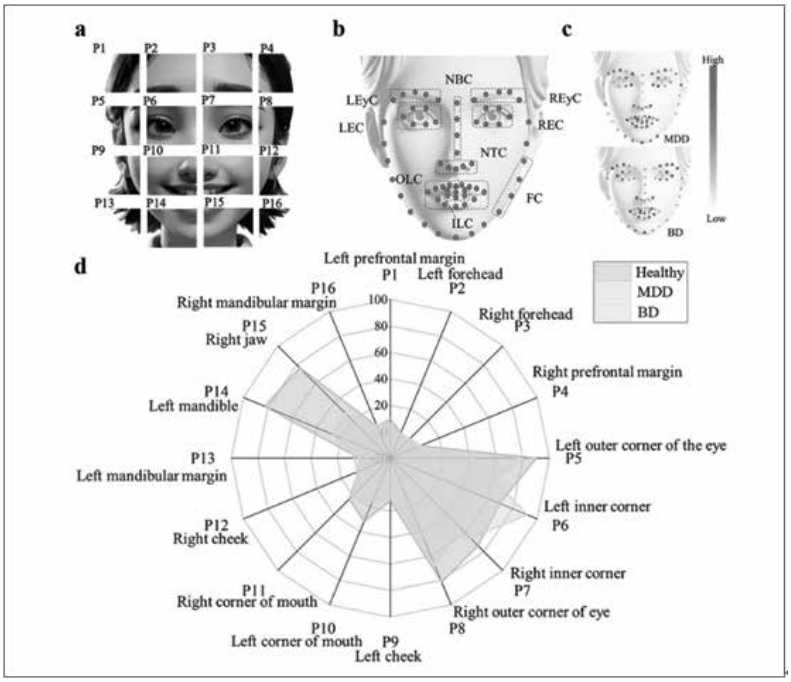

图1-14

2. 感性工学

感性工学是一门工程学分支技术，其核心在于将用户对产品的主观感受与需求进行量化，并探寻这些感性变量与工程物理参数之间的关联，最终将这些关系转化为具体的设计形态。该方法的主要优势在于，它能够基于理性视角，采用定量或半定量方式来剖析感性意象的应用原理，从而将原本模糊的感性体验转化为精确的工程数据。这不仅有助于人们

更准确地表达自身需求，也为设计提供了客观依据。在其研究过程中，“以人为本”是贯穿始终的首要理念，这使其成果能更有效地服务于提升人们的生活品质。

感性工学的标准研究流程通常包含以下4个步骤。

(1) 收集数据：通过调研或仪器测量，掌握消费者对特定产品的心理感受或生理反应。

(2) 分析数据：运用统计学、工程学等方法，系统分析所收集的感性数据。

(3) 转译数据：对分析结果进行解读，并将其映射、应用到新产品的设计领域中。

(4) 实现设计：最终完成一款承载了目标感性意象的新产品设计，如图1-15所示。

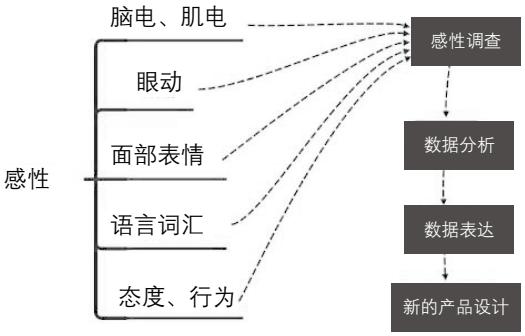

图1-15

感性工学在情感智能中的应用，核心在于将人的主观感受、情绪等“感性”因素，通过科学方法量化并融入机器的“智能”系统，从而构建更能理解和适应人的产品与服务。简单来说，感性工学就像一位 “情绪翻译官” 或 “感受测量师”。它通过系统的研究方法，把用户那些模糊、感性的体验(如“这个设计让我感到舒适安心”“这个交互让我有点焦虑”)，转化为工程师和AI能够理解和处理的具体设计参数与数据模型。

表1-14展示了感性工学赋能情感智能的几个主要方向。

表1-14

方向	感性工学的应用	典型应用
情感化交互设计	分析用户对不同界面、语音语调、交互流程的情感反馈，量化出“友好”“高效”“可信赖”等感性意象对应的最佳设计参数	智能车载系统根据分析，采用让多数司机感到平静、专注的界面配色与语音交互逻辑
个性化产品推荐与服务	建立用户情感偏好模型，不仅分析用户“买了什么”，更理解产品带来的情感体验(如“治愈”“奢华”)，实现从“物”到“感受”的推荐升级	音乐、视频流媒体不仅根据历史记录，更根据用户当前情绪(通过分析表情、语气等)推荐契合心境的歌单或影片
情感计算与识别辅助	为情感识别算法提供可解释的、工程化的“感性标签”，帮助AI更准确地关联生理信号(如表情、心率)与具体的内在情感状态	在心理健康领域，帮助AI系统更精准地区分“悲伤”和“平静”的细微面部特征差异(见图1-14所示案例)

(续表)

方向	感性工学的应用	典型应用
智能产品创新	在产品开发前期，系统性研究目标用户群对新产品概念的感性期待，将“温暖”“科技感”等需求直接转化为材料、形态、功能的指导	设计一款旨在缓解孤独感的陪伴机器人时，将其外观、触感、互动方式设定为“友好、非侵入性”等具体参数

感性工学的引入，使情感智能超越了单纯的情绪识别，迈向深度的情感理解与主动的体验塑造。它正推动AI从“理解人的指令”向“理解人的感受”进化，是创造真正“以人为中心”的下一代智能系统的关键技术桥梁。

1.3.8 类脑智能

人工智能的运行也极其耗能。根据国际能源署的估计，2022年，全球数据中心、加密货币和人工智能的总耗电量已高达约460 TWh(太瓦时)，几乎占到全球总电力需求的2%。更值得关注的是，该机构预测到2026年，这一数字将翻倍增长，超过1000 TWh。

如今，人工智能正面临算力需求与能源消耗之间的根本矛盾：模型的性能越强大，其能耗规模也呈指数级增长。由成千上万张GPU组成的计算集群，如同一个个巨大的“电力怪兽”，吞噬着海量能源。与此同时，驱动这些“怪兽”的“燃料”——用于模型训练的高质量数据，正被预测将面临枯竭，这构成了更深层的资源约束。

在此背景下，类脑智能正受到日益广泛的关注，并被视为一条极具前景的新兴路径。它旨在模拟人脑的神经结构、信息处理方式与认知策略，从而有望在自主学习能力和超低能耗两方面实现突破，展现出推动人工智能范式变革的巨大潜力。

同时，类脑智能(Brain-Inspired Intelligence)远不止于对人脑神经网络的简单模拟，其核心在于从神经科学与认知科学中汲取根本原理。它通过模拟人脑的神经结构、信息处理路径和认知策略，旨在构建在自主学习能力、能效、泛化性与自适应性方面都更为优秀的新一代智能系统。在研究范式上，类脑智能可大致分为以下两类。

(1) 结构启发：即模仿大脑神经元、突触与神经回路的拓扑与动力学特征，形成如卷积神经网络(CNN)、递归神经网络(RNN)、深度信念网络(DBN)与脉冲神经网络(SNN)等模型，这类模型试图通过类脑的计算机制增强感知与表征能力。

(2) 行为启发：注重模仿人类在学习、注意、记忆、决策与迁移等方面的认知过程，从而发展出注意力机制、自监督学习、强化学习、模仿学习、迁移学习、机器反遗忘等方法，这些机制显著提升了AI系统对复杂环境与任务的应对能力。

现今在AI领域被广泛应用的许多实践，都源于神经科学研究提供的理论基础。例如：

- 视觉皮层的层级编码直接启发了卷积神经网络的设计。
- 海马体的空间导航功能促进了机器人拓扑地图建模的发展。
- 前额叶皮层在决策与记忆中的作用，激发了AI对顺序推理与时间依赖建模的尝试。
- 大脑的冗余信息处理与极致能耗控制，为低功耗神经计算提供了理论方向。

1.4　人工智能的关键技术

人工智能的关键技术主要包括机器学习、知识图谱、自然语言处理、计算机视觉、语音识别、生物特征识别以及人机智能交互等。其中，机器学习作为核心驱动力，使系统能够从数据中自主学习并持续优化；知识图谱构建了结构化的知识基础，支持机器进行逻辑推理与深度理解；自然语言处理与计算机视觉分别赋予AI处理文本和图像信息的感知能力，而语音识别与生物特征识别则进一步扩展了其在听觉感知与身份认证方面的应用；人机智能交互技术致力于实现自然、高效的人机沟通；机器人技术则综合集成上述各项技术，为AI在物理世界中的具身行动与任务执行提供能力支撑。

1.4.1　机器学习

机器学习不仅是人工智能的核心技术支柱，更是实现各类智能能力的关键路径。作为一门多学科交叉的领域，它融合了统计学、计算机科学、优化理论、神经科学等多个学科的理论与方法，致力于研究计算机如何模拟人类学习行为，从而获取新知识、形成新技能，并基于经验持续优化自身性能。

基于数据的机器学习是现代智能技术的重要代表，其核心思想是从已知的观测数据(样本)中寻找内在规律和模式，并利用所学规律对未知数据或未来情形进行有效预测与推断。这一方法构成了当前人工智能实现与应用的基石(本书第3章将详细介绍机器学习技术)。

1.4.2　知识图谱

知识图谱作为人工智能时代的关键技术之一，通过结构化的方式有效地表示现实世界中的实体、属性及其关系，从而实现了数据向知识的高效转化。其本质是语义网络与图结构的结合，利用符号表示和嵌入技术构建多维度的关联知识库，为复杂场景提供决策支持，如图1-16所示。

1. 什么是知识图谱

想象一下，我们脑子里的知识是什么样子的？是像一本厚厚的词典，所有词条按A~Z的顺序排列，彼此孤立地躺在那里吗？例如以下信息。

- 姚明：中国篮球运动员，身高2.26米。
- NBA：美国职业篮球联赛。
- 休斯顿火箭队：NBA球队。

这样记忆信息，不仅效率低下，更关键的是，它完全割裂了信息之间本应存在的千丝万缕的联系。而知识图谱，正是为了解决这个问题而生的。它就像一位超级智能的“知识整理大师”，其核心任务就是为世界上的信息编织一张清晰、互联的关系网络。

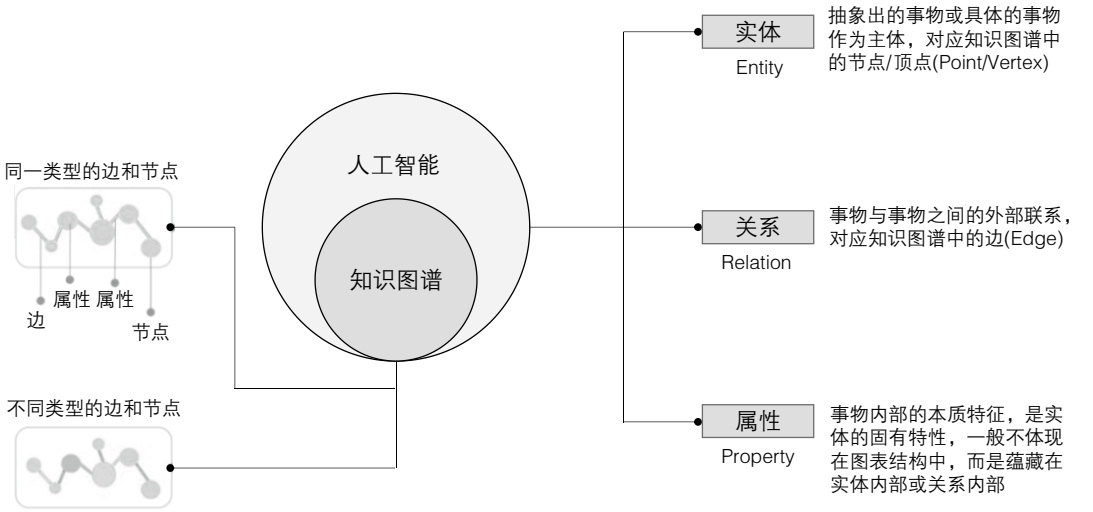

图1-16

在这张关系网中：

- 节点就像一个个“小泡泡”，代表具体的实体。例如，姚明、NBA、休斯顿火箭队、中国、篮球。
- 边就像连接泡泡的“线”，代表实体间的关系。例如：
  - 姚明→效力于→休斯顿火箭队。
  - 休斯顿火箭队→属于→NBA。
  - 姚明→国籍是→中国。

信息结构瞬间变得清晰、立体。姚明不再是一个孤立的名词，而是通过“效力于”“国籍是”等关系，与休斯顿火箭队、NBA(美国职业篮球联赛)、中国等实体动态连接，共同构成一张生动的知识网络。

这张由节点和边编织而成的、互联互通的知识网络，就是知识图谱的核心形态。图1-17所示为一张简单的知识图谱。

2. 知识图谱的构建

简单来说，知识图谱=实体+关系+属性(例如，实体姚明具有“身高2.26米”这一属性)。它的核心作用在于，通过“关系”将零散、孤立的知识点，编织成一张巨大的、互联互通的“知识地图”。

这张“知识地图”并非凭空产生，而是由人工智能通过一套系统化的流程构建而成，其核心原理可分为三步，如表1-15所示。

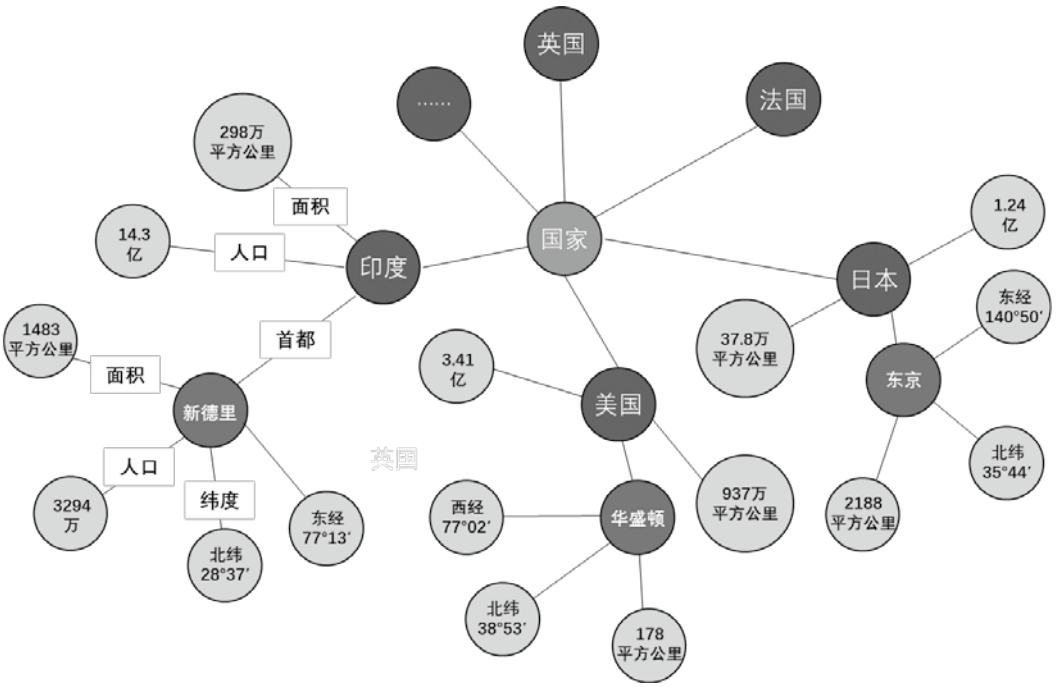

图1-17

表1-15

步骤	核心任务	AI工作	举例
1. 实体识别	识别文本中的关键对象(找“泡泡”)	类似信息采集员，从网页、新闻、数据库等多源数据中，通过自然语言处理技术识别并提取出人名、地名、组织名等实体	在句子“姚明在休斯顿火箭队打篮球”中，能识别出姚明(人名)、休斯顿火箭队(组织)、篮球(事物)3个实体
2. 关系抽取	确定实体之间的关联(画“连线”)	分析实体所在的上下文，抽取出它们之间的语义关系，将孤立的实体连接起来	从同一句中，分析出“姚明”和“休斯顿火箭队”之间存在“在...打”的动作，可将其提炼为“效力于”的关系
3. 知识融合与存储	整合信息并结构化存储(织成“大网”)	对不同来源的同一实体(如“姚明”和“姚主席”)进行消歧与合并，将清洗后的结构化关系网络存入图数据库	将关于“姚明”的所有信息(如效力球队、国籍)融合到同一节点下，最终形成可查询的、互联的知识图谱

构建知识图谱的整个过程，就如同AI在进行一场宏大的信息拼图：它将互联网上散落、无序的信息碎片，通过识别、关联与整合，最终拼接成一幅结构化、可理解的世界知识图景。

3. 知识图谱的作用

知识图谱这张“知识地图”的用途非常广泛。它使人工智能从机械的“信息检索” 迈向深度的“知识理解与推理”，从而变得更“智能”。其主要价值体现在以下几个方面。

(1) 智能搜索(从“找到信息”到“理解信息”)：传统搜索仅返回包含关键词的零散链接。而基于知识图谱的搜索，能直接生成整合的“知识卡片”，清晰展示实体(如人物)的结构化信息及其关联。例如搜索“姚明”，结果会主动归纳其生平、社会关系等。这背后是谷歌、百度等巨头正在应用的核心技术。

(2) 精准推荐(从“行为反馈”到“兴趣推理”)：传统推荐依赖于用户的历史行为(如观看记录)。知识图谱则通过实体关联(如NBA→篮球→运动装备)，构建用户的“兴趣图谱”，实现跨领域的联想推荐。例如，从用户喜欢NBA，可能关联推荐篮球鞋或相关纪录片，而不仅是更多比赛视频。

(3) 复杂问答(从“关键词匹配”到“关系推理”)：传统问答难以处理需要多步逻辑的问题。知识图谱赋予AI沿关系链进行推理求解的能力。例如，回答“姚明的前队友现在在哪个队当教练？”，AI可沿“效力于→队友→现任教练”等关系路径逐步推导出答案，展现出类似人类的思考过程。

(4) 认知基石(从“处理数据”到“理解世界”)：知识图谱为AI提供了结构化的世界知识模型和关系推理框架。这使其不仅能处理数据，更能理解数据背后的语义与联系，是机器掌握常识、进行复杂决策，并最终迈向通用人工智能的关键基础。

综上所述，知识图谱就如同为人工智能装配了一张超级知识地图，使其能够将零散、孤立的信息点，转化为一个相互关联、可被推理的智慧网络。

1.4.3 人机智能交互

人机交互(Human-Computer Interaction，HCI)是一门研究用户与系统之间交互关系的学科。这里的“系统”范围很广，可以指各种机器、计算机化的软件或复杂的硬件设备。

人机交互的核心在于用户可见并与之直接沟通的交互界面。用户正是通过这个界面与系统“对话”并下达指令。它无处不在，小到收音机的播放按键，大到飞机的仪表板乃至发电厂的控制中心。优秀的人机交互界面设计，其核心在于匹配用户对系统的心智模型。这意味着设计必须充分考虑用户如何理解和预期系统的运作，其根本目标是实现系统的高可用性与用户体验的友好性。

人类社会正从计算机时代迈向人工智能时代，在此过程中，人机关系经历了深刻演变，由以操作为中心的人机交互，逐步转向以协作为核心的人机智能交互(Human-Computer Intelligent Interaction，HCII)。HCII在传统人机交互的基础上，深度融合了自然语言处理、机器学习等前沿人工智能技术，致力于系统性地探索并构建更加高效、自然的人与智能系统之间的新型交互范式。

在传统计算机科学中，研究重点往往集中于如何让计算机更高效地执行任务；而在

人机智能交互领域，关注的核心则转向如何实现人与计算机之间更自然、更流畅的交流与协作。这种交互方式不仅需要理解人的认知模式和行为习惯，还需融合机器学习、深度学习、自然语言处理等先进技术，共同构建智能化的交互体验。

图1-18所示为人机智能交互中人工智能与人机交互的关系。

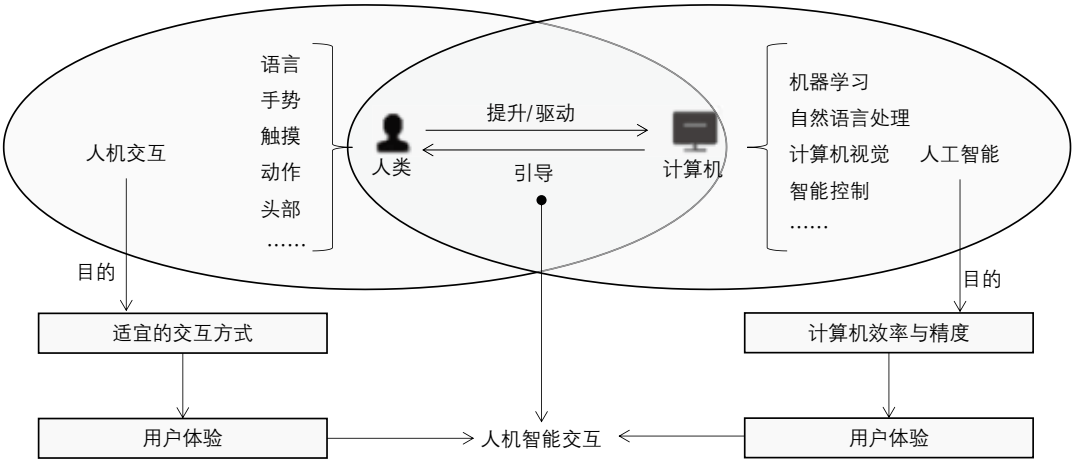

图1-18

在数字化浪潮中，设计师的职责已不再局限于创造美观的界面，更延伸至如何让界面具备智能、实现与人自然流畅的交流。在开展人机智能交互设计时，理解其应用架构尤为关键。图1-19所示的人机智能交互五层应用架构，可作为我们学习过程中的重要理论参照。

(1) 传感器层(Sensor Layer)：该层是系统的“感官”，负责从环境中收集原始数据，主要包括麦克风(捕捉声音信号)、摄像头(获取图像与视频)、眼动追踪仪(追踪用户视线移动)、各类环境传感器(监测温度、光照等物理参数)。

(2) 信号处理层(Signal Processing Layer)：这一层可视为系统的“感知系统”，主要负责将传感器获取的原始信号转化为计算机可处理的信息，典型处理包括语音识别(将音频转换为文字)、人脸识别(检测与识别人脸)、手势识别(解析人体动作)、图像处理(分析与增强视觉内容)。

(3) 智能决策层(Intelligent Decision-making Layer)：该层担任系统的“大脑”，负责理解用户意图并生成决策，其核心支撑技术包括机器学习(从数据中提取规律)、深度学习(处理更复杂的模式识别问题)、自然语言处理(理解与生成人类语言)。

(4) 用户界面层(User Interface Layer)：用户直接与之交互的界面层(该层是交互设计师关注的核心部分)，其设计直接影响使用体验，主要包括以下交互形式：图形界面(如按钮、图标等视觉元素)、语音界面(支持语音指令交互)、手势界面(通过动作进行控制)。

(5) 应用层(Application Layer)：该层体现为面向最终用户的完整应用场景，例如智能家居系统(实现对灯光、温度等设备的智能控制)、智能交通系统(包括自动驾驶、自适应信号灯等)。

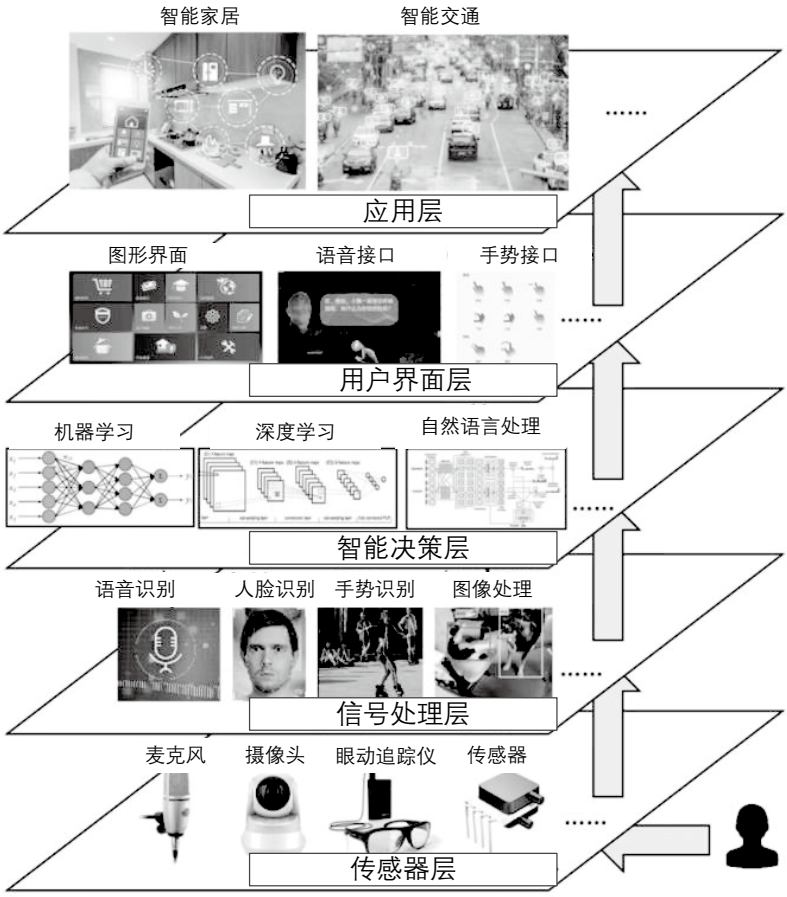

图1-19

过去五年间，人机智能交互领域的研究主要聚焦于数据处理、脑机接口、多模态情绪识别、动作识别以及语言处理等方向(感兴趣的用户可以扫描右侧二维码进一步了解相关知识)。

人机智能交互的主要研究领域

1.4.4 自然语言处理

自然语言处理(Natural Language Processing，NLP)作为人工智能领域的关键技术，致力于使计算机能够理解、分析并生成人类日常使用的自然语言(如中文、英文、西班牙语等)。其研究与实践不仅关注文本的表层信息，更注重捕捉语境、语义与情感等深层含义。例如，系统需能辨别“你真厉害(赞扬)”与“你可真厉害(反讽)”之间的微妙差别，亦需理解“他借了我一本书”中“借”这一动作的语义指向。只有达成这样的深度理解，计算机才算真正“读懂”人类语言。

举个简单的例子：当人类说出“打开窗户”时，人们能够立刻理解其中的意图——操作的对象是窗户，执行的动作是打开。然而对于计算机而言，这串文字起初只是无意义的字符序列。它需要通过自然语言处理技术，先将“打开窗户”分解为“打开(动词)”与“窗户(名词)”，进而识别出“动作与对象”之间的逻辑关系，最终将其转换为机器能够执行

的指令。可以说，自然语言处理如同一位“人机语言翻译官”，既接收人类传递的语言信息，又将其翻译为计算机可理解的逻辑形式，从而搭建起人机之间高效沟通的桥梁(本书后面的章节中将进一步介绍自然语言处理技术)。

1.4.5　计算机视觉

计算机视觉(Computer Vision，CV)是一个致力于通过技术手段使计算机能够“看见”并“理解”图像内容的研究领域。例如，它使得计算机能够解析图片和视频中的信息，识别其中的对象、场景与行为，从而实现对视觉世界的智能化认知。

1. 什么是计算机视觉

计算机视觉是人工智能领域的关键技术之一，其核心目标是使计算机能够“看懂”图像或视频中的内容。举例来说，它需要能够判断一张照片中的宠物是猫还是狗、识别图像中的人物是老张还是老王，甚至分析视频里的人物正在做什么。

从技术定义上讲，计算机视觉是指利用摄像设备与计算机系统，替代人眼实现对目标的识别、跟踪与测量，并进一步对图像进行处理，使其更便于人眼观察或适合仪器检测。作为一门科学学科，计算机视觉致力于研究相关理论与技术，构建能够从图像或多维数据中提取高层次信息的人工智能系统。从工程视角看，它旨在开发可模仿人类视觉功能的自动化系统，以完成各类视觉任务。

计算机视觉的远期理想，是使计算机能像人类一样通过视觉感知来理解世界，并具备适应环境自主决策的能力。然而，真正实现这一目标仍面临巨大挑战：尽管摄像机拍摄的图像与人眼所见的内容相同，但对计算机而言，任何图像都仅仅是像素值的排列组合，是一组没有生命的数字。如何让计算机从这些“冰冷”的数字中提取出有意义的视觉信息，正是计算机视觉所要解决的根本问题。

2. 计算机视觉的原理

用过相机或手机的人都知道，计算机在拍摄图像方面具有出色的保真度和细节还原能力。甚至可以说，其人工“视觉”在信息捕捉的精度上已经超越了人类与生俱来的视觉系统。但正如“听见不等于听懂”一样，“看见”也并不意味着“看懂”，要让计算机真正理解图像内容，仍然是一个具有挑战性的任务。

从本质上看，图像是由像素组成的网格，每个像素的颜色由红、绿、蓝三种基色的强度组合而成，即RGB值。通过调节这三种颜色的配比，我们几乎可以表示出任何一种颜色。

最简单且适合入门的计算机视觉算法之一是颜色跟踪，例如追踪一个粉色小球。具体步骤是：先记录小球中心像素的RGB值作为目标颜色，然后将图像输入程序，让程序找出与目标颜色最接近的像素。算法可以从图像的左上角开始，逐个计算每个像素颜色与目标颜色之间的差异。在遍历所有像素后，差异最小的那部分像素很可能就对应小球所在的位置。该算法不仅可用于单张图像，也能逐帧应用于视频中，实现对运动目标的持续跟踪。当然，由于光线、阴影等环境因素的干扰，小球的实际颜色会与存储的RGB值有所偏差，

但仍能保持较高的相似度。然而在极端条件下，例如夜间进行足球比赛，跟踪效果可能会显著下降；如果某队球衣颜色与球的颜色相近，算法甚至可能完全失效。因此，除非环境条件得到严格控制，这类基于颜色的跟踪算法在实际应用中有较大局限。

目前，主流的计算机视觉算法普遍采用以卷积神经网络为代表的深度学习技术。由于其出色的性能，CNN已成为该领域应用最广泛的方法之一。鉴于深度学习知识体系较为广泛，本节不做详细展开，相关内容可参阅本书第4章。

3. 计算机视觉的任务

计算机视觉的典型任务有以下几项。

(1) 图像分类：图像分类是指根据图像的语义信息对不同类别的图像进行区分，是计算机视觉的核心技术，也是物体检测、图像分割、物体跟踪、行为分析、人脸识别等其他高层次视觉任务的基础。例如，如图1-20所示，通过图像分类，计算机能够识别出图像中的人(person)、树(tree)、草地(grass)和天空(sky)。

图像分类在许多领域都有广泛的应用，包括安防领域的人脸识别和智能视频分析，交通领域的交通场景识别，互联网领域的基于内容的图像检索和相册自动归类，以及医学领域的图像识别等。

(2) 目标检测：目标检测的任务是给定一张图像或者一个视频帧，让计算机找出其中所有目标的位置，并给出每个目标的具体类别。如图1-21所示，以识别和检测人为例，用边框标记图像中所有人的位置。

图1-20

图1-21

然而，在多类别目标检测中，一般使用不同颜色的边框对检测到的不同物体的位置进行标记，如图1-22所示。

(3) 语义分割：语义分割是计算机视觉中的一项基础任务，其目的是将视觉输入划分为不同的语义可解释类别。在语义分割过程中，我们将整幅图像分割成像素组，并对这些像素组进行标记和分类。例如，我们可能需要区分图像中所有属于汽车的像素，并将这些像素涂成蓝色。如图1-23所示，图像中的元素被标记为人(红色)、树木(深绿色)、草地(浅绿色)和天空(蓝色)。

图1-22

图1-23

(4) 目标追踪：目标跟踪是指对图像序列中的运动目标进行检测、提取、识别和跟踪的过程。通过这一过程，我们可以获取运动目标的运动参数，并对其进行处理与分析，从而实现对运动目标行为的理解，以便完成更高级的检测任务。

计算机视觉作为人工智能领域中发展最为迅速、应用最为广泛的技术分支之一，被誉为人工智能的“眼睛”，能够帮助各行各业捕捉、解析并挖掘出更深层的信息。伴随着算法持续迭代、硬件算力不断提升、数据规模爆发式增长，以及5G高速网络的全面赋能，计算机视觉的应用前景必将迎来更加广阔的发展空间。

1.4.6　生物特征识别

生物特征识别是人工智能的重要研究方向，属于感知智能的范畴，其主要目标是让机器具备准确区分不同个体的能力。该技术基于每个个体独一无二的生理或行为特征，实现对其身份的识别与认证。

生物特征识别技术的典型应用流程通常包括注册和识别两个阶段。

- 注册阶段首先通过传感器采集人体生物信息，例如利用图像传感器获取指纹、人脸等光学信息，借助麦克风录制说话声等声学信息。采集到的数据经过预处理与特征提取，生成相应的生物特征模板并存入数据库。
- 识别阶段则采用与注册阶段相同的采集方式，对识别个体进行信息采集、预处理与特征提取，随后将提取的特征与数据库中存储的模板进行比对分析，从而完成身份识别。

从应用任务来看，生物特征识别主要分为辨认与确认两类。辨认是指从特征库中确定待识别个体身份的过程，属于“一对多”的搜索问题；确认则是将待识别个体的生物特征与库中某一特定模板进行比对，以验证其身份，属于“一对一”的匹配问题。

生物特征识别涵盖的技术种类广泛，包括指纹、掌纹、人脸、虹膜、指静脉、声纹、步态等多种模态。其实现过程涉及图像处理、计算机视觉、语音识别、机器学习等多种技术。目前，生物特征识别已成为智能化身份认证的关键技术，在金融、公共安全、教育、交通等领域得到广泛应用。下面将对指纹识别、人脸识别、虹膜识别、指静脉识别、步态识别及声纹识别等主流技术，分别进行介绍。

(1) 指纹识别：指纹识别过程通常包括数据采集、数据处理和分析判别3个步骤。首先，在数据采集阶段，系统利用光学、电容、热敏或压力等物理传感器获取指纹图像。随后，在数据处理阶段，对采集到的图像进行预处理、畸变校正和特征提取等操作。最后，在分析判别阶段，系统对提取的特征进行比对与判别，从而完成身份验证，如图1-24所示。

(2) 人脸识别：人脸识别是计算机视觉中的典型应用。从应用过程来看，人脸识别技术可以分为3个主要步骤：检测与定位、面部特征提取与比对、人脸确认，如图1-25所示。然而，人脸识别技术的应用受到多个因素的影响，包括光照、拍摄角度、图像遮挡和年龄等。在受限条件下，人脸识别技术已相对成熟，而在自由环境下，相关技术仍在不断改进与发展。

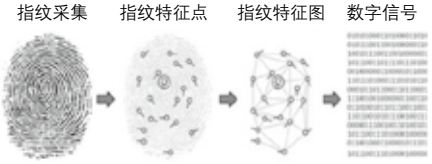

图1-24

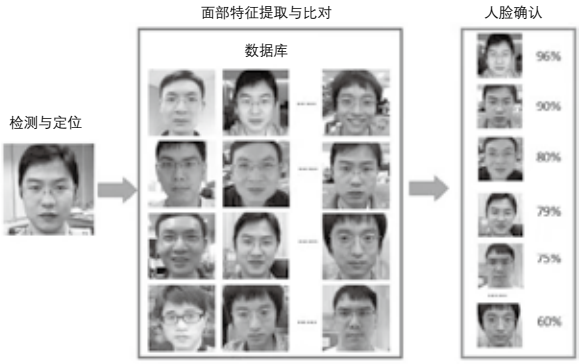

图1-25

(3) 虹膜识别：虹膜识别的理论框架主要包括虹膜采集、虹膜定位、虹膜图像增强、虹膜区域归一化处理、特征提取和识别，如图1-26所示。大多数研究工作都是基于这一理论框架展开的。在虹膜识别技术的应用中，主要面临两个挑战：传感器性能和光照影响。

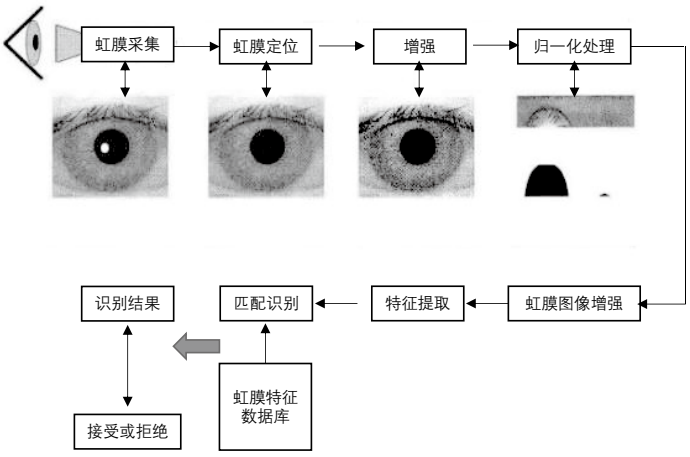

图1-26

首先，由于虹膜的尺寸较小且受到黑色素的遮挡，必须在近红外光源下使用高分辨率图像传感器才能获得清晰的成像，这对传感器的质量和稳定性提出了较高的要求。其次，光照强度的变化会导致瞳孔的缩放，从而引发虹膜纹理的复杂形变，增加了与虹膜匹配的难度。

(4) 指静脉识别：指静脉识别是一种利用人体静脉血管中脱氧血红蛋白对特定波长范围内的近红外线具有良好吸收特性的技术，通过近红外光对指静脉进行成像与识别，如图1-27所示。由于指静脉血管的分布具有很强的随机性，其网络特征具备良好的唯一性，并且作为人体内部特征，不易受到外界因素的影响，因此模态特性相当稳定。然而，指静脉识别技术在应用过程中面临的挑战主要源于成像单元的性能。

(5) 步态识别：步态是远距离复杂场景下唯一可清晰成像的生物特征，步态识别是指通过身体体型和行走姿态来识别人的身份，如图1-28所示。相比上述几种生物特征识别，步态识别的技术难度更大，体现在其需要从视频中提取运动特征，以及需要更高要求的预处理算法，但步态识别具有远距离、跨角度、光照不敏感等优势。

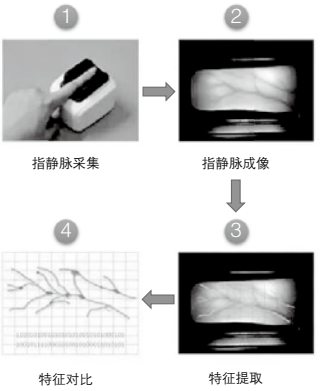

图1-27

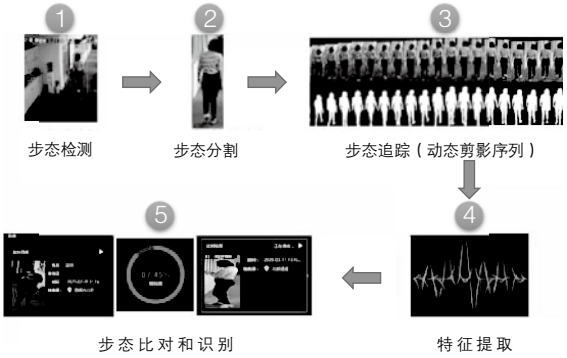

图1-28

(6) 声纹识别：声纹识别是一种根据待识别语音的声纹特征来识别说话人的技术。声纹识别通常分为前端处理和建模分析两个阶段。在识别过程中，首先对某段来自特定说话者的语音进行特征提取，然后将提取的特征与多种声纹模型库中的声纹模型进行匹配。常用的识别方法包括模板匹配法和概率模型法等。

1.5　本章小结

本章系统性地介绍了人工智能的基本概念、发展脉络与技术体系。首先，从“人类行为与理性思维”等维度探讨了人工智能的多重定义，并梳理了其自达特茅斯会议以来的三次发展浪潮与生成式AI的最新进展。其次，按照研究内容将人工智能划分为计算、感知、认知、行为、群体、混合、情感及类脑智能等层次，并详细阐述了其按功能、能力、学习方式等多种维度的分类体系。最后，重点剖析了机器学习、知识图谱、自然语言处理、计算机视觉、生物特征识别等关键技术的原理与应用，以及人机智能交互等前沿方向，为初学者构建了人工智能领域的整体知识框架。

1.6 习题

1. 简述人工智能从“人类与理性”“思想与行为”两个维度衍生出的四种定义类型，并各举一例。

2. 按照能力等级分类，人工智能可分为哪三个层次？请各举一项具有代表性的技术。

3. 人工智能经历了哪三次发展浪潮？第三次浪潮的主要技术特征是什么？

4. 知识图谱的构建主要包含哪三个核心步骤？它在人工智能应用中主要发挥哪些作用？

5. 生物特征识别一般分为哪两种应用任务？请简要说明两者的区别。