



计算机类专业“互联网+”创新型精品教材

人工智能应用导论

主 编 杨 静 侯艳芳 袁正道

北京出版集团
北京出版社

人工智能应用导论

主 编 杨 静 侯艳芳 袁正道

北京出版集团
北京出版社

图书在版编目 (CIP) 数据

人工智能应用导论 / 杨静, 侯艳芳, 袁正道主编

—北京: 北京出版社, 2023.6

ISBN 978-7-200-17970-5

I . ①人… II . ①杨… ②侯… ③袁… III . ①人工智能—高等教育—教材 IV . ① TP18

中国国家版本馆 CIP 数据核字 (2023) 第 101241 号

人工智能应用导论

RENGONG ZHINENG YINGYONG DAOLUN

主 编: 杨 静 侯艳芳 袁正道

出 版: 北京出版集团

北 京 出 版 社

地 址: 北京北三环中路 6 号

邮 编: 100120

网 址: www.bph.com.cn

总 发 行: 北京出版集团

经 销: 新华书店

印 刷: 定州启航印刷有限公司

版 印 次: 2023 年 6 月第 1 版 2023 年 6 月第 1 次印刷

成品尺寸: 185 毫米 × 260 毫米

印 张: 15

字 数: 337 千字

书 号: ISBN 978-7-200-17970-5

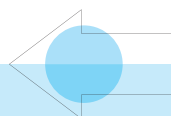
定 价: 48.00 元

教材意见建议接收方式: 010-58572162 邮箱: jiaocai@bphg.com.cn

如有印装质量问题, 由本社负责调换

质量监督电话: 010-82685218 010-58572162 010-58572393

项目一 认识人工智能——揭开人工智能的面纱	1
任务一 人工智能知识认知	2
任务二 人工智能技术发展历程探究	6
任务三 探究新职业——人工智能训练师	11
任务四 人工智能应用支撑认知	27
项目二 驾驭物体识别——让机器更聪明	46
任务一 植物检测	47
任务二 探究鸢尾花品种	70
任务三 识别图像中的动物	81
任务四 手写数字识别	85
项目三 语音识别技术应用——让机器对你言听计从	98
任务一 语音数据采集	99
任务二 语音转文字	107
任务三 语音合成	118
任务四 聊天机器人	125
任务五 校园智能客服	133
项目四 人脸识别技术应用——机器也认识你	152
任务一 人脸检测	153
任务二 人脸特征提取	162

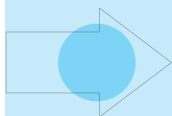


任务三	人脸识别	171
任务四	照片智能搜索	175
任务五	口罩检测赋能卫生防护	184

项目五 多姿多彩的生活——人工智能应用秀..... 194

任务一	人物动漫化	195
任务二	智能音乐创作	203
任务三	智能驾驶	208
任务四	我知道你想买什么	215

参考文献..... 232



项目一

认识人工智能——揭开人工智能的面纱

人工智能自诞生以来，在短短几十年的时间里取得了巨大的进展，并在多个领域得到了广泛的应用，极大地改变了人们的社会生活。从无人工厂到无人驾驶，从智慧城市到智能娱乐，从能跳舞又能送盒饭的机器人到手机里的指纹解锁、人脸识别、语音输入等，人工智能已经从象牙塔飞入寻常百姓家，推进了生产，提供了便捷的生活方式。那么人工智能是什么呢？让我们一起来揭开它的面纱。

知识目标

1. 了解人工智能的定义。
2. 理解现代人工智能及分类。
3. 了解人工智能发展历程。
4. 了解我国人工智能发展现状和战略规划。
5. 知道人工智能训练师工作流程及内容。
6. 了解人工智能技术的三大基础支撑：算力、算法、数据。

技能目标

1. 会搜集信息、辨别信息。
2. 能制作符合模型训练要求的图像数据集。
3. 会选择合适的算法模型进行操作。
4. 能使用工具进行数据标注。

素养目标

1. 树立文化自信，增强民族自豪感。
2. 培养学生正确面对曲折和困难的人生观。
3. 培养学生科技自立自强意识。
4. 训练辩证思维能力和发展观。

任务一

人工智能知识认知

任务描述

人工智能正在逐步地融入我们的生产生活中，那么，人们常说的人工智能到底是什么呢？本任务让我们和李强一起来学习人工智能的概念、人工智能的三个方面以及人工智能的未来发展等知识。

任务目标

1. 能从各类人工智能定义中总结自己对人工智能的认识。
2. 能发现生活中的人工智能应用。



人工智能的作用

任务实施

一、人工智能的概念

人工智能诞生于 1956 年，至今已有半个多世纪的发展历程。这样算起来人工智能并不是一项新技术，但是作为现在最前沿的交叉学科，学界尚未对人工智能有统一的定义。不同领域对于人工智能有着不同的理解，所以不同领域对于人工智能的定义也不同。

(1) 新华词典在线版里的人工智能。

人工智能是计算机科学的一个分支，研究应用计算机来模拟人类的某些智力活动，从而代替人类的某些脑力劳动。人工智能是一门涉及数学、心理学、生物学、语言学、经济学、哲学和法律学等方面的综合性学科，主要研究模式识别、学习过程、探索过程、推理过程等。

(2) 《人工智能标准化白皮书（2018 年）》里的人工智能。

人工智能是利用数字计算机或者由数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

(3) 人工智能之父眼里的人工智能。

人工智能之父——约翰·麦卡锡对人工智能的定义是“制造智能机器的科学与工程，特别是智能计算机程序”。

(4) 研究领域的人工智能。

人工智能是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术学科，它是计算机科学的一个分支。

(5) 应用领域的“人工智能”。

人工智能是一门综合学科，主旨是研究和开发出智能实体，在这一点上它属于工程

学。那么人工智能自然包括工程学的一些基础学科，还包括数学、逻辑学、归纳学、统计学、系统学、控制学、工程学、计算机科学，且包括对哲学、心理学、生物学、神经科学、认知科学、仿生学、经济学、语言学等其他学科的研究，可以说这是一个集数门学科精华的尖端学科中的尖端学科。



人工智能（AI）这个概念是1956年达特茅斯会议上正式提出来的。到底什么是人工智能呢？虽然已经经历了几十年的发展，但学界、业界尚未有统一的定义。不同领域不同类型的专家学者有不同的观点，目前归纳起来可以从两个维度来认识人工智能：一是思考，二是行动。

二、人工智能的三个方面

在人工智能的发展过程中，不同学科背景的人工智能学者对它有着不同的理解。综合起来，我们可以从能力、学科和实用三个方面对人工智能进行定义。从能力的角度来看，人工智能是指用人工的方法在机器上实现的智能，即计算智能；从学科的角度来看，人工智能是研究如何构造智能机器或智能系统，使它能模拟、延伸和扩展人类智能的学科，即感知智能；从实用的角度来看，人工智能是指用机器实现所有目前必须借助人类智慧才能实现的任任务，即认知智能。

（一）计算智能

计算智能是指机器可以像人类一样存储、计算和传递信息，帮助人类存储和快速处理海量数据，有利于算法的优化和硬件的技术进步。这一阶段是感知智能和认知智能的基础。

（二）感知智能

感知智能是指机器具有类似人的感知能力，如视觉、听觉等，不仅可以听懂、看懂，还可以基于此做出判断并做出反馈或采取行动，即“能听会说，能看会认”。目前研究较多、成果显著的感知智能技术包括图像识别、语音识别等，国内外人工智能技术发展均集中于这一阶段。

（三）认知智能

认知智能是指机器能够像人一样主动思考并采取行动，全面辅助或替代人类工作，是人工智能的最高级形态，也是行业未来的着力点。

三、人工智能的未来

人工智能在未来是否会拥有意识，我们现在还不得而知。但是，很多研发人员和工程师希望人工智能可以拥有意识。

数字克隆人是人工智能在未来的发展方向之一。数字克隆人是人类个体的思维方式和兴趣爱好的数字复制品，但是，数字克隆人的开发可能只是一种用于实现人格的技

术。2015 年左右,人们已经开发出用于实现此目标的传感技术。由此实现了很多事情,例如,根据图像来推测面部表情,并将其与情绪关联等。今后利用传感技术模拟人格的尝试会越来越多。

知识链接

人工智能从最初的“玩简单游戏”(跳棋等)到现在的“玩复杂游戏”(围棋、星际争霸等),经历了各种风风雨雨。现在,你甚至不需要问人工智能可以干什么,你细心看看周围就可以发现,人工智能已经应用于我们生活的方方面面,如高铁站的人脸识别验票进站、手机的语音助手等。

一、工业

现阶段,工业企业在智能制造趋势下纷纷开始探索智能化转型的路径,基于工业大数据分析的工业智能蕴藏着巨大商业价值的革命性技术,越来越多地受到企业青睐。例如:使用人工智能技术,通过工艺控制与设备知识模型实现工艺参数优化、协同生产流程优化;通过图像检测算法辅助工人对缺陷定位和分类,有效控制质量异常,减少人力成本;通过对关键的设备运行参数进行建模,判断机器的运行状态,预测维护时间;通过对生产制程工艺参数建模来预测产品指标、推动生产优化、提升良品率等。利用这些技术能有效地提高工业水平,加快工业发展,更有效地为人们服务。

二、商业

电商兴起以后,商业已经变成了一个数据驱动的行业,最典型的就是亚马逊商城。亚马逊商城的人工智能推荐系统是一个强大的引擎,可以为你推荐各种你可能购买的商品。

随着游戏市场竞争的日趋激烈,越来越多的游戏运营服务选择借助大数据挖掘出更多更细的用户群来进行精细化、个性化运营,从而更好地抓住用户,获得更高的收益。在游戏运营中,无论是流失挽留,还是拉新,以及付费用户预测都是游戏运营的重要内容。

目前,人工智能在商业中得到了广泛的应用,其包括智能客服机器人、决策支持、预测营销、语音搜索、图像识别等,相信人工智能在商业中的应用将有着良好的发展前景。

三、金融

金融本来就是数字的游戏,不过现在正从数字(number)转向数据(data)。金融机构通过数据分析客户信用,进行诈骗检测、证券交易等。

近年来,国外先进大型银行均不同程度地受到金融危机的波及,对大型长期内部项目的持续投入捉襟见肘。而当前,建行以企业级姿态推动利用互联网理念思维实现各项传统业务转型,正是向互联网龙头企业学习借鉴最先进的理念技术、构建“建行大脑”、争取摘取“王冠上的宝石”的最佳时机。“建行大脑”通过将建行经营管理工作进行全方位数字化和自动化,运用人工智能技术,提升银行经营效率和客户体验,无论从外部客户角度,还是从内部员工角度来看,建设银行就好像拥有一个“无所不知的大脑”。

四、医疗

在医学领域，开始人工智能实践的首先是涉及图像的B超、CT、病理专业等，其次是内镜诊断领域。医学影像是疾病诊断的主要路径之一，因此，通过机器读取医学影像成为一个热点，无数的科研工作者已经对此展开了广泛的研究。

人工智能可以在医疗行业多个流程中发挥价值，问诊、医疗报告生成、初步筛查、医疗影像的辅助诊断等，通过大规模的医疗历史数据进行训练，在多个流程中发挥价值，也可以缓解国内医疗资源不均、不足等问题。

2023年3月，微软旗下的临床文档软件公司Nuance在其最新的语音转录应用程序中添加了GPT4；4月，微软和Epic宣布将把OpenAI的GPT-4引入医疗保健领域，以帮助医护人员回复患者信息和分析医疗记录。同月，谷歌宣布将向用户群发布其医学大模型Med-PaLM 2。国内方面，卫宁健康、科大讯飞、创业慧康等积极布局，行业应用加速探索。

人工智能在其他领域也有各种各样的应用，这里就不再展开说明了。相信随着5G时代的到来，会有更多的设备利用网络获取更多人们所需要的数据，出现更丰富的应用。



知识拓展

新职业档案：人工智能训练师

人工智能训练师是指使用智能训练软件，在人工智能产品实际使用过程中进行数据库管理、算法参数设置、人机交互设计、性能测试跟踪及其他辅助作业的人员。

人工智能训练师在2020年正式成为新职业并纳入国家职业分类目录。随着2021年底该职业的国家职业技能标准颁布，业内开展职业培训和人才技能评价有了基本依据。人社部发布的报告显示，经测算我国人工智能人才缺口超过500万。

主要工作任务：

1. 标注和加工图片、文字、语音等业务的原始数据；
2. 分析提炼专业领域特征，训练和评测人工智能产品相关算法、功能和性能；
3. 设计人工智能产品的交互流程和应用解决方案；
4. 监控、分析、管理人工智能产品应用数据；
5. 调整、优化人工智能产品参数和配置。

思考与练习

一、填空题

1. 人工智能之父_____对人工智能的定义是“制造智能机器的科学与工程，特别是_____”。
2. 人工智能是一门_____学科，主旨是研究和开发出_____，在这一

点上它属于_____。

3. 从能力的角度来看,人工智能是指用人工的方法在机器上实现的智能,即_____;从学科的角度来看,人工智能是研究如何构造智能机器或智能系统,使它模拟、延伸和扩展人类智能的学科,即_____;从实用的角度来看,人工智能是指用机器实现所有目前必须借助人类智慧才能实现的任务,即_____。

4. _____是人类个体的思维方式和兴趣爱好的数字复制品,但是,_____的开发可能只是一种用于实现人格的技术。

二、判断题

1. 人工智能是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术学科,它是人工技术科学的一个分支。()

2. 人工智能还包括心理学、生物学、经济学等其他学科的研究。()

3. 认知智能是指机器能够像人一样主动思考并采取行动,全面辅助或替代人类工作。()

4. 计算智能是感知智能和认知智能的基础。()

三、思考题

说说“人工智能技术”给你的生活带来的改变。

任务二

人工智能技术发展历程探究

任务描述

作为人工智能时代的学生,我们在享受人工智能带来的“福利”的同时,也要了解人工智能技术的发展历程。本任务我们紧随李强的学习步伐来了解人工智能技术起起落落螺旋式发展的历程。

任务目标

能从人工智能技术发展历程中总结其发展规律。

任务实施



人工智能的发展方向

一、人工智能的第一次高峰

1950年,一名在校大四学生马文·明斯基与他的同学邓恩·埃德蒙建造了世界上第一台神经网络计算机,这被视为人工智能的起点,马文·明斯基也被世人称为“人工智能之父”。同年,“计算机之父”阿兰·图灵提出了一个举世瞩目的想法,即图灵测试。图灵当时大胆预言了真正具备智能的机器的可行性。1956年,在达特茅斯学院举

办的一次会议上，计算机专家约翰·麦卡锡提出了“人工智能”一词。这次会议后不久，约翰·麦卡锡和马文·明斯基先后搬到了麻省理工学院（Massachusetts Institute of Technology）并一起开展人工智能相关研究，两人共同创建了世界上第一个人工智能实验室——MIT AI LAB 实验室。人们在达特茅斯会议上正式确立了人工智能这一术语，对人工智能的学术研究也正式拉开序幕，人工智能学者和技术开发者陆续涌现，成为第一批人工智能领域的研究实践者。在此后的 10 多年里，计算机被广泛应用于数学和自然语言领域，用来解决代数、几何和英语等问题，这大大增加了研究学者对机器人智能化发展的信心，不少研究者甚至认为：“20 年内，机器将能完成人能做到的一切。”人工智能迎来第一次高峰。

二、人工智能的第一次低谷

20 世纪 70 年代，人工智能进入了一段痛苦而艰难的岁月。科研人员在人工智能的研究中对项目难度预估不足，导致与美国国防高级研究计划署的合作计划失败，使人工智能的前景蒙上了一层阴影。与此同时，社会舆论的压力也慢慢压向人工智能这边，导致很多研究经费被转移到了其他项目上。当时，人工智能面临三个技术瓶颈：①计算机性能不足。这导致早期很多程序无法在人工智能领域得到应用。②问题的复杂性高。早期人工智能程序主要是解决特定的问题，因为特定的问题对象少，复杂性低，可一旦问题的维度上升，程序立刻就不堪重负了。③数据量严重不足。在当时不可能找到足够大的数据库来支撑程序进行深度学习，这很容易导致机器无法读取足够量的数据从而进行智能化。人工智能发展第一次进入低谷。

三、人工智能的崛起

1980 年，卡内基梅隆大学为数字设备公司设计了一套名为 XCON 的“专家系统”。这是一种采用人工智能程序的系统，是具有完整专业知识和经验的计算机智能系统，可以简单地将其理解为“知识库 + 推理机”的组合。这套系统在 1986 年之前能为公司每年节省下来超过 4000 万美元的经费。这种商业模式衍生出了像 Symbolics、Lisp Machines 等和 IntelliCorp、Aion 等硬件、软件公司。在这个时期，仅专家系统产业的价值就高达 5 亿美元，人工智能应用再次进入繁荣期。

四、人工智能的第二次低谷

1987 年，苹果和 IBM 公司生产的台式机性能都超过了 Symbolics 等厂商生产的专家系统通用计算机，从此，专家系统风光不再，曾经轰动一时的人工智能系统就此宣告结束其历史进程，人工智能应用再次进入低谷期。

五、人工智能再次崛起

20 世纪 90 年代中期，随着 AI 技术尤其是神经网络技术的逐步发展，以及人们对 AI 开始抱有客观理性的认知，人工智能技术开始进入平稳发展时期。1997 年 5 月 11 日，IBM 的计算机系统“深蓝”战胜了国际象棋世界冠军卡斯帕罗夫，又一次在公众领域引发了现象级的 AI 话题讨论，成为人工智能发展的一个重要里程碑。2006 年，

Hinton 在神经网络的深度学习领域取得突破，人类又一次看到机器赶超人类的希望，这是人工智能标志性的技术进步。2009 年，IBM 首席执行官彭明盛首次提出“智慧地球”这一概念。“智慧地球”战略的主要内容是把新一代 IT 技术充分运用到各行各业之中，即把感应器嵌入和装备到电网、铁路、桥梁、隧道、公路、建筑、供水系统、大坝、油气管道等各种物体中，实现万物互联，形成“物联网”的物理系统，然后基于超级计算机和云计算技术实现与物理系统的交互，从而帮助人类更加精细和动态地管理生产和生活，最终达到“智慧”状态。

知识链接

我国人工智能研究不但起步较晚，而且发展道路曲折坎坷，遭受过质疑、批评，甚至被打压，可谓举步维艰，但是改革开放之后，我国人工智能开启了快速发展之路。我国人工智能经历了科研起步和产业快速发展阶段，现在已经进入国家战略规划发展阶段。

一、科研起步阶段

1978 年 3 月，邓小平发表了“科学技术是生产力”的重要讲话，提出“向科学技术现代化进军”的战略决策。这一决策开启了解放思想的先河，促进了中国科学事业的发展，使中国科技事业迎来了科学的春天。广大科技人员思想大解放，酝酿着解禁人工智能研究。吴文俊凭借几何定理的机器证明成果，获得了 1978 年全国科学大会重大科技成果奖，成为国际自动推理界的领军人物，他所开创的数学机械化也在国际上被誉为“吴方法”。

20 世纪 70 年代末至 20 世纪 80 年代前期，一些人工智能相关项目已被纳入国家科研计划。例如，在 1978 年召开的中国自动化学会年会上，展示了光学文字识别系统、手写体数字识别、生物控制论和模糊集合等研究成果，表明中国人工智能在生物控制和模式识别等方向的研究已开始起步。1986 年起，智能计算机系统、智能机器人和智能信息处理等重大项目被列入国家高技术研究发展计划（“863 计划”）。人工智能研究已经成为国家重点支持的科研领域。

二、产业快速发展阶段

2003 年，西安电子科技大学雷达信号处理国家重点实验室和北京大学智能科学系共同提出成立智能科学与技术专业。智能科学与技术面向前沿高新技术的基础性本科专业，覆盖面很广。智能科学与技术专业涉及机器人技术，以新一代网络计算为基础的智能系统，微机电系统（MEMS），与国民经济、工业生产及日常生活密切相关的各类智能技术与系统，新一代的人—机系统技术等。全国共有 36 所本科院校开展了智能科学与技术专业的人才培养，为我国人工智能从科研实验进入产业应用提供了源源不断的支持，我国人工智能开始进入快速发展期。浪潮天梭在 2006 年 8 月以 3 胜 5 平 2 负击败柳大华等 5 位中国象棋大师组成的联盟。科大讯飞语音识别技术已经处于国际领先地位，其语音识别和理解的准确率均达到了世界第一，自 2006 年首次参加国际权威的 Blizzard Challenge 大赛以来，一直保持冠军地位。百度推出了度秘和智能驾驶汽车。腾

讯推出了机器人记者 Dreamwriter 和图像识别产品腾讯优图。阿里巴巴推出了人工智能平台 DTPAI 和机器人客服平台。清华大学成功研发出的人脸识别系统以及智能问答技术都已经获得了应用。中科院自动化所成功研发了“寒武纪”芯片并建成了类脑智能研究平台。华为也推出了 MoKA 人工智能系统。

三、国家战略规划发展阶段

世界各国已经认识到人工智能是未来国家之间竞争的关键，因而纷纷开始部署人工智能发展战略，以期占领新一轮科技革命的历史高点，我国也不例外。2017 年 3 月，人工智能首次被写入政府工作报告。2017 年 7 月，国务院发布《新一代人工智能发展规划》，提出了“三步走”战略目标：第一步，到 2020 年，人工智能总体技术和应用与世界先进水平同步，人工智能产业成为新的重要经济增长点，核心产业规模超过 1500 亿元，带动相关产业规模超过 1 万亿元；第二步，到 2025 年，新一代人工智能在智能制造、智能医疗、智慧城市、智能农业、国防建设等领域得到广泛应用，核心产业规模超过 4000 亿元，相关产业规模超过 5 万亿元；第三步，到 2030 年，人工智能理论、技术与应用总体达到世界领先水平，形成涵盖核心技术、关键系统、支撑平台和智能应用的完备产业链和高端产业群，人工智能核心产业规模超过 1 万亿元，带动相关产业规模超过 10 万亿元。2017 年 10 月，人工智能被写入党的十九大报告。2018 年 3 月，人工智能再度被写入政府工作报告。2018 年 12 月的中央经济工作会议上，人工智能被列入“新基建”的核心板块。2019 年 3 月，人工智能第三次出现在政府工作报告中，并升级为“人工智能+”。2020 年 3 月，中共中央政治局常务委员会召开会议，提出要发力于科技端的基础设施建设，人工智能再次成为“新基建”七大板块中的重要一项。

新技术推动学科建设和催生新职业。2018 年 4 月，教育部在研究制定《高等学校引领人工智能创新行动计划》时确定设立人工智能专业，进一步完善中国高校人工智能学科体系，旨在培养中国人工智能产业的应用型人才，推动人工智能一级学科建设。2019 年 3 月，全国共有 35 所高校获首批人工智能新专业建设资格。教育部确定 2019 年度增补人工智能技术服务专业，自 2020 年起执行，首批共有 171 所高职院校获批新专业人才培养资格。2019 年 4 月，中华人民共和国人力资源和社会保障部（以下简称人社部）等部门发布 13 个新职业，包括人工智能工程技术人员等。2020 年 2 月，人社部再次向社会发布了未来急需的 16 个新职业，人工智能训练师、智能制造工程技术人员等名列其中。

在人工智能领域，无论是从理论研究、技术研发方面，还是从产业基础方面来看，应该说我国的研究积累与发达国家相比差距不大，目前很多方面已经处于世界领先水平。人工智能已经成为我国战略规划部署的重点，未来应用发展空间无限。

四、国际上人工智能领域的发展现状

人工智能是引领新一轮科技革命和产业变革的战略性技术，在培育新兴产业、促进传统产业升级、优化产业结构、推动生产力整体跃升等方面具有重要作用，正成为世界各国密切关注并竞相发展的热点领域。

（一）资源层：数据及算力的大幅提升将 AI 推向浪潮之巅

互联网及数字经济的快速发展，为 AI 算法训练提供充裕的数据来源。人工智能领域顶级专家吴恩达曾提到：发展人工智能就像用火箭发射卫星，需要强大的引擎和足够的燃料，算法模型就是其引擎，高性能的算力是打造引擎的工具，海量的数据就是引擎的燃料。基于深度学习的算法特点，其计算准确度与数据量基本成正比。当前，机器学习所能应用的数据已经不局限于文本、数字等结构化数据，还包括视频、音频、图片等非结构化数据。而根据 IDC（互联网数据中心）的统计数据，2021 年全球数据量已经达到 82ZB，预计到 2026 年将达到 214 ZB。

（二）技术环境：开源框架大幅降低开发门槛

国际各巨头公司在 2016 年纷纷创立 AI 开源的开发框架，人工智能的开发框架是底层硬件与上层软件之间的纽带，可以视作人工智能进行开发和应用的“操作系统”。以往的专家系统是基于本地化专业知识进行运算的，以知识库和推理机为中心进行展开，推理机设计内容由不同的专家系统应用环境决定，不具备通用性。同时，知识库是开发者收集录入的专家分析模型与案例的资源集合，只能在单机系统环境下使用且无法连接网络，升级更新不便。企业的软件框架实现有闭源和开源两种形式，少数企业选择闭源方式开发软件框架，目的是打造技术壁垒。目前，业内主流软件框架基本上基于开源化运营，如谷歌的 TensorFlow、脸书的 Torchnet、微软的 DMTK、IBM 的 SystemML、三星的 VELES 等，均具有分布式深度学习数据库和商业级即插即用功能。

（三）外部环境：政策助力，人工智能迎来发展良机

当前，全球人工智能领导地位争夺赛已进入白热化阶段。近年来，英国、法国、德国、加拿大、日本、印度、新加坡等国陆续发布了人工智能战略，积极推动人工智能研究开发和产业应用。随着现代科学技术的迅猛发展，机器的算力、算法和海量场景数据等方面获得重大突破，使人工智能重获新生，呈现出井喷爆发之势。虽然美国因最早运用大数据、云计算、物联网等先进技术在数字经济领域占据领导地位，但是美国要在人工智能领域中保持领导力，还需要根据自身的政治经济军事的发展需要而制定相应的人工智能战略。中国密集出台人工智能相关政策，并将 AI 上升至国家战略，力求在下一轮工业革命中抢占先机。自 2015 年以来，国内不断出台推动人工智能发展的鼓励政策：建立人工智能促进机制；为人工智能提供支持，如税收优惠、财政扶持及制定金融政策等；推动创新，包括促进产业集群，布局创新基地，鼓励人才培养；政府制定规制与保障措施；等等。



知识拓展

人工智能云服务

人工智能云服务，一般也被称为 AIaaS（AI as a Service，AI 即服务）。这是目前主流的人工智能平台的服务方式。具体来说，AIaaS 平台会把几类常见的 AI 服务进行拆

分，并在云端提供独立或打包的服务。这种服务模式类似于开了一个 AI 主题商城，所有的开发者都可以通过 API 使用平台提供的一种或多种人工智能服务，部分资深的开发者还可以使用平台提供的 AI 框架和 AI 基础设施来部署与运维自己专属的机器人。

国内典型的例子有腾讯云、阿里云和百度云。以腾讯云为例，目前该平台提供 25 种不同类型的人工智能服务，其中有 8 种偏重场景的应用服务，15 种侧重平台的服务，2 种能够支持多种算法的机器学习和深度学习框架。

思考与练习

一、填空题

1. 邓小平发表了“_____”的重要讲话，提出“向科学技术现代化进军”的战略决策。
2. 新技术推动学科建设和_____。
3. 我国人工智能经历了_____和_____，现在已经进入国家战略规划发展阶段。
4. 20 世纪 90 年代中期，随着 AI 技术尤其是_____的逐步发展，以及人们对 AI 开始抱有客观理性的认知，人工智能技术开始进入_____发展时期。

二、判断题

1. “智慧地球”战略实现了万物互联，形成“物联网”的物理系统。()
2. 人工智能已经被写入党的十九大报告。()
3. 吴文俊开创的数学机械化也在国际上被誉为“吴方法”。()
4. 在人工智能领域，我国的研究积累与发达国家相比差距不大，目前很多方面已经处于世界领先水平。()

三、思考题

搜集人工智能近 70 年发生各个重大事件时期我国的社会、政治、经济的发展，谈谈你对各学科之间关联的理解。

任务三

探究新职业——人工智能训练师

任务描述

随着对人工智能了解的深入，李强知道了有一个新兴的职业——人工智能训练师。本任务就是我们跟着李强一起走近人工智能训练师，在百度 EasyDL 平台实践体验新职业的工作流程，如图 1-1 所示，完成图像分类的数据处理，制作数据集，进行模型训练和模型校验，最后在公有云完成部署并生成 H5 二维码，如图 1-2 所示。

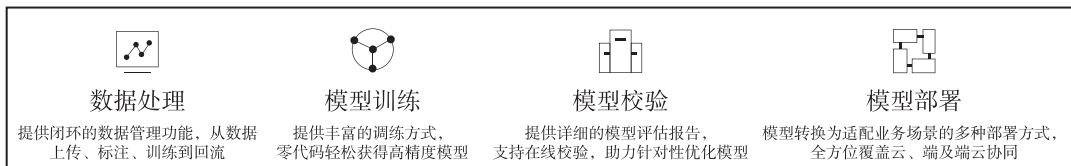


图 1-1 EasyDL 人工智能模型训练流程



图 1-2 图像分类模型应用 H5 体验程序

任务目标

1. 能完成图像标注。
2. 能制作符合模型训练要求的图像数据集。
3. 能利用平台完成模型训练。

任务实施

随着 AI 落地的深入，不少企业发现，在越来越多的实际应用场景中，需要结合场景数据进行模型的定制。有研究显示，这样的定制化需求占比高达 86%。但与之相矛盾的是，大部分中小企业并不具备专业的算法开发能力，导致开发定制 AI 模型对于企业来说难以实现。针对这种情况，百度、华为、腾讯、阿里等 AI 头部企业都提供了创建模型—上传并标注数据—训练并校验模型效果—模型部署的一站式 AI 服务，实现了全流程自动化，用户只需根据平台的提示进行操作即可，使人工智能零基础的用户也能快速实现人工智能应用。

下面以百度深度学习定制 EasyDL 平台来体验人工智能训练师从数据标注到模型部署的整个工作流程。EasyDL 平台根据目标客户的应用场景及深度学习的技术方向，

开放了 EasyDL 图像、EasyDL 文本、EasyDL 语音、EasyDL OCR、EasyDL 视频以及 EasyDL 结构化数据六个模型类型。本次任务我们选择 EasyDL 图像，通过训练简易的图像分类模型，实现猫和狗的分类。这一过程将通过创建图像分类模型、上传数据、标注数据、训练模型、校验和发布五个步骤进行，在该过程中不追求模型的最终效果，而是着重于基本操作。

一、创建图像分类模型

(1) 打开 EasyDL 图像官网，进入 EasyDL 图像页面，如图 1-3 所示。



图 1-3 EasyDL 图像页面

(2) 单击图 1-3 中的“立即使用”按钮，在弹出来的“选择模型类型”对话框中选择“图像分类”选项，如图 1-4 所示，进入登录界面，输入账号和密码。未注册的用户需要先注册。



图 1-4 “选择模型类型”对话框

(3) 进入图像分类模型管理界面，在左侧的导航栏中，选择“我的模型”标签页，如图 1-5 所示，单击“创建模型”按钮，进入信息填写界面。

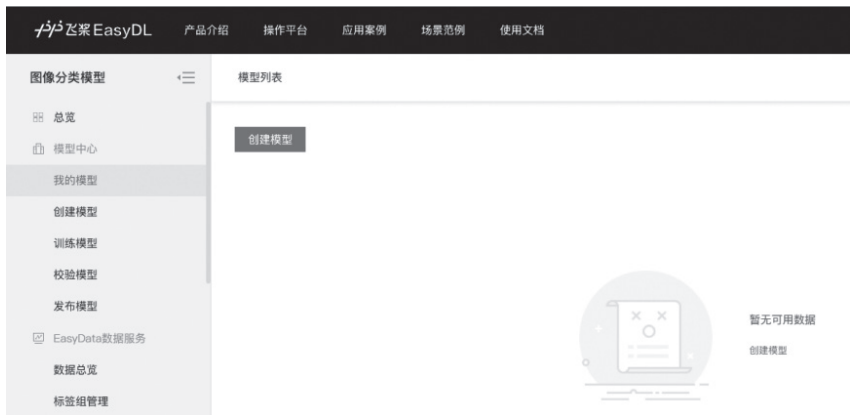


图 1-5 “我的模型”标签页

(4) 在信息填写页面中填入“模型名称”“邮箱地址”“联系方式”和“业务描述”，选择“模型归属”，所有带有星号的选项都必须填写，完成效果如图 1-6 所示。



图 1-6 “创建模型”页面

(5) 信息填写完成后，单击“完成创建”按钮即可创建成功。在左侧导航栏“我的模型”标签页中即可看到所创建的模型，如图 1-7 所示。



图 1-7 创建的模型列表

二、上传数据

(1) 选择左侧导航栏中的“数据总览”标签页，如图 1-8 所示，单击“创建数据集”按钮，进入创建数据集信息填写页面。



图 1-8 数据总览页面

(2) 按照提示填写信息。在“数据集名称”一栏可输入你自己取的数据集名称，其他均为默认项，无需修改，如图 1-9 所示。信息填写完成后，单击“完成”按钮。



图 1-9 创建数据集页面

(3) 数据集创建成功后，在界面中将出现该模型的数据集信息，包括版本、数据集ID、数据量、标注类型、标注状态、清洗状态等，如图 1-10 所示。

创建数据集									
猫和狗		数据集ID: 487004		新增版本 全部版本 删除					
版本	数据集ID	数据量	最近导入状态	标注类型	标注模板	标注状态	清洗状态	操作	
V1	1756366	0	● 已完成	图像分类	单图单标签	0% (0/0)	-	多人标注	导入 删除 共享

图 1-10 数据集列表

(4) 单击图 1-10 中的“导入”链接，进入导入数据页面，在“数据标注状态”一栏选中“无标注信息”单选按钮，在“导入方式”一栏选择“本地导入”和“上传压缩包”选项，如图 1-11 所示。单击“上传压缩包”按钮，进入上传压缩包页面，如图

1-12 所示。单击“已阅读并上传”按钮，选择本地保存的压缩数据文件，如图 1-13 所示。单击“打开”按钮开始上传文件，上传成功后，如图 1-14 所示，单击“确认并返回”按钮。



图 1-11 导入数据页面



图 1-12 上传压缩包页面

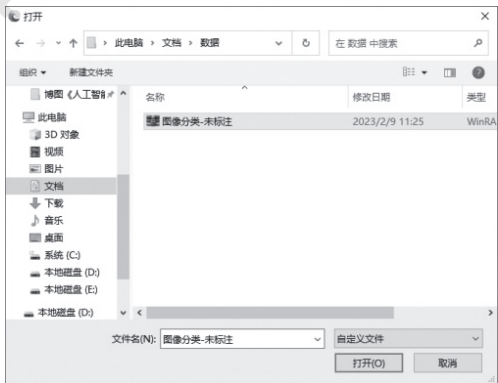


图 1-13 上传压缩文件



图 1-14 压缩文件上传成功

(5) 此时数据开始导入，数据导入状态如图 1-15 所示，等待数据全部导入。



图 1-15 数据导入状态

三、标注数据

(1) 数据集导入完成后，可以看到最近导入状态已更新为“已完成”，数据量、标注状态操作都有变化，如图 1-16 所示。单击该数据集右侧操作栏下的“查看与标注”按钮，进入标注界面。



图 1-16 导入数据集后的数据列表

(2) 添加标签“猫”和“狗”，单击“添加标签”按钮，如图 1-17 所示，输入“猫”，如图 1-18 所示，单击“确定”按钮，标签就添加成功了，效果如图 1-19 所示。利用相同的方法添加一个“狗”标签，完成后如图 1-20 所示。

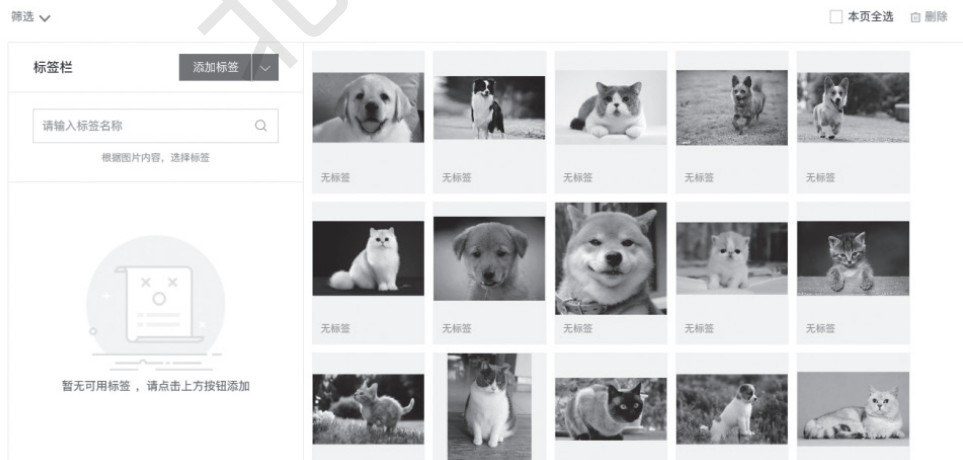


图 1-17 添加标签页面

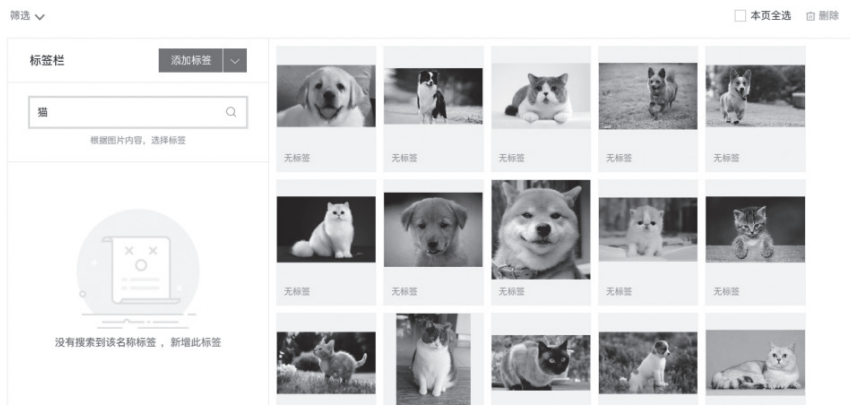


图 1-18 添加“猫”标签

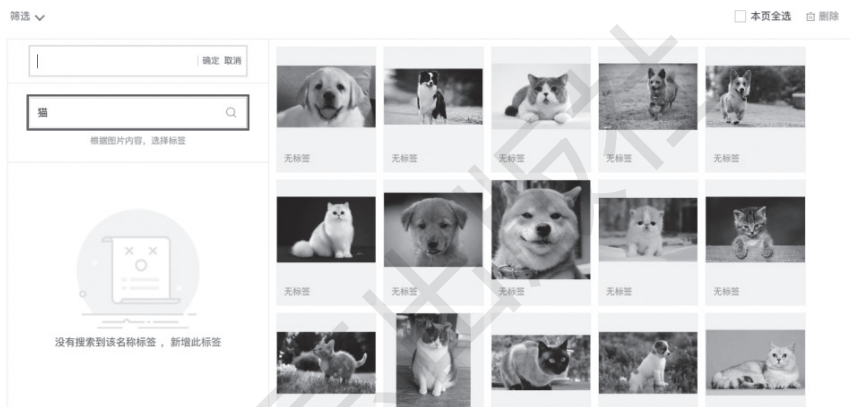


图 1-19 添加“猫”标签成功

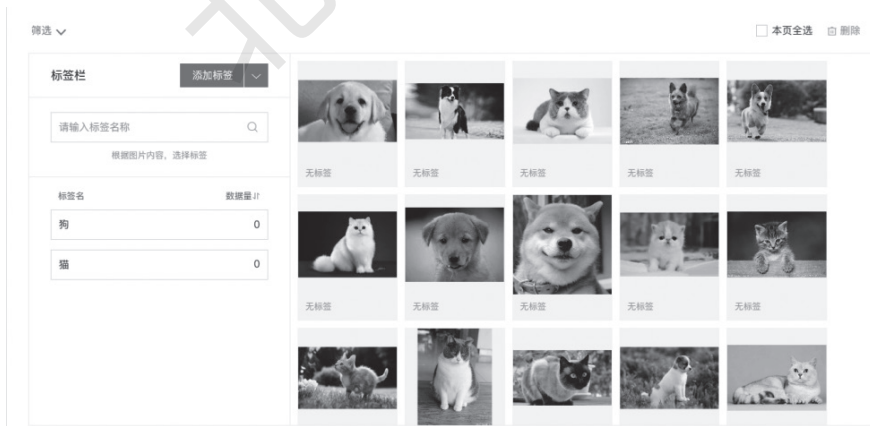


图 1-20 添加“猫”“狗”标签成功

(3) 标签添加完成后, 即可进行数据标注。单击界面右上方的“批量标注”按钮, 如图 1-21 所示, 进入标注界面。

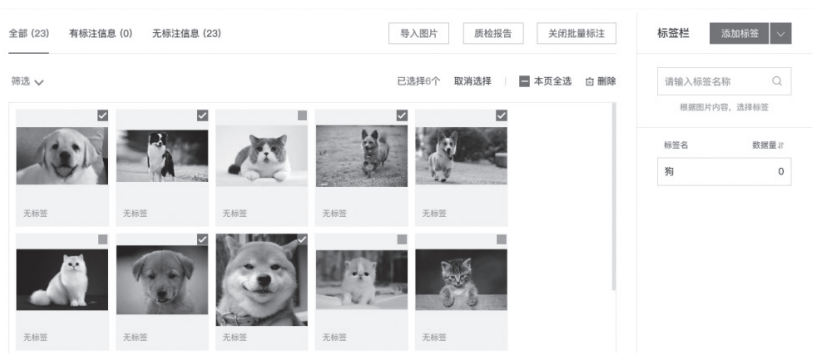


图 1-21 选择批量标注

(4) 选中所有代表狗的图像，单击右侧标签栏下对应的标签名“狗”，如图 1-22 所示，即可进行标注。此处要注意不可选错图像，若标注错误，会严重影响模型的训练效果。



图 1-22 批量标注“狗”的图像

(5) 利用相同的方法完成“猫”的图像的标注，全部标注完成后的效果，如图 1-23 所示。接下来回到数据总览页面，标注状态如图 1-24 所示。

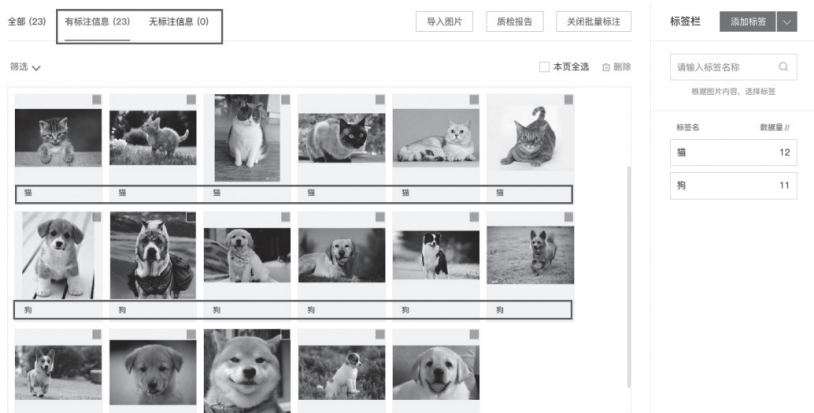


图 1-23 标注完成状态

版本	数据集ID	数据量	最近导入状态	标注类型	标注模板	标注状态	清洗状态	操作
V1	1756366	23	● 已完成	图像分类	单图单标签	100% (23/23)	-	查看与标注 多人标注 导入 清洗 ...

图 1-24 数据标注完成后的数据列表

四、训练模型

(1) 选择模型训练，进入训练模型页面，分别选择模型、选择算法、添加数据集，如图 1-25 所示。单击“添加数据集”→“请选择”进入添加数据集页面，如图 1-26 所示。单击图 1-26 中的“+”号进行数据添加，添加猫、狗两类数据集的效果如图 1-27 所示。



图 1-25 训练模型页面



图 1-26 添加数据集页面



图 1-27 添加猫狗数据集

(2) 添加完数据集，训练模型页面的添加数据集效果如图 1-28 所示，其他参数取默认值，然后单击“开始训练”按钮。



图 1-28 添加数据集后的训练模型页面

(3) 在弹出的对话框中，单击“继续训练”按钮，即可继续进行训练，如图 1-29 所示。

(4) 单击“训练状态”→“训练中”旁的感叹号图标，可查看训练进度，如图 1-30 所示，还可以设置在模型训练完成后，发送短信至个人手机号。训练时长与数据量大小有关，本次训练大约耗时 3~15 分钟，训练完成后的效果如图 1-31 所示。

提示



有些标签的训练的文件 少于20个，是否添加更多的数据以获得更好的训练效果？

放弃训练

继续训练

图 1-29 单击“继续训练”按钮



图 1-30 模型“训练中”状态



图 1-31 模型“训练完成”状态

(5) 在版本配置界面, 可以查看该训练任务的开始时间、任务时长、训练时长及训练算法等基本信息。在“训练数据集”一栏, 可以查看各分类的训练效果。查看模型评估报告, 如图 1-32 所示。



图 1-32 模型评估报告页面

五、校验和发布

(1) 单击图 1-31 中的“校验”链接, 进入校验模型页面, 单击“启动校验”按钮, 再单击页面中间的“点击添加图片”按钮, 选择事先准备好的一张非数据集中的图像, 等待校验, 如图 1-33 所示。



图 1-33 校验模型页面

(2) 校验完成后, 在页面中可以看到模型的识别结果, 在页面右侧可以查看预测分类及其对应的置信度, 如图 1-34 所示。

(3) 选择“发布模型”进入发布模型页面, 填写服务名称和接口地址, 完成后单击“提交申请”按钮后进入发布状态, 如图 1-35 所示。

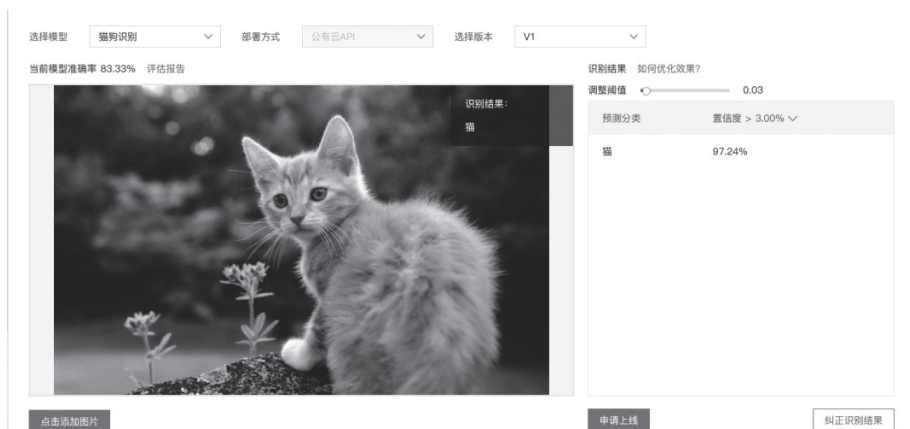


图 1-34 校验结果页面



图 1-35 发布模型页面

(4) 回到我的模型列表中, 模型服务状态为“发布中”, 如图 1-36 所示。

服务名称	模型名称	模型版本	服务状态	调用单价 ①	更新时间	操作
ceshi	猫狗识别	V1	发布中	-	---	服务详情 更新版本 体验H5

图 1-36 模型“发布中”状态

(5) 模型发布成功后, 服务状态为“已发布”, 同时操作页面中会增加“体验 H5”链接, 如图 1-37 所示。单击“体验 H5”链接, 进入页面, 如图 1-38 所示, 选择调用 App, 单击“下一步”按钮进入“自定义样式”页面, 输入信息, 如图 1-39 所示, 完成后单击“生成 H5”按钮。

发布新服务

控制台

输入模型名称

Q

服务名称	模型名称	模型版本	服务状态	调用单价 ①	更新时间	操作
ceshi	猫狗识别	V1	已发布	-	--	服务详情 更新版本 体验H5

图 1-37 模型“已发布”状态



图 1-38 体验 H5 页面



图 1-39 “自定义样式”页面

(6) 进入“体验 H5”的“完成”页面，如图 1-40 所示。此时即可拿起手机，扫二维码进行测试，测试效果如图 1-41 所示。



图 1-40 “体验 H5”的“完成”页面



图 1-41 测试效果

知识链接

人工智能技术的发展使我们的生产生活发生了巨大的变化，人工智能训练师是近年来随着 AI 技术广泛应用而产生的新职业，2020 年 2 月，人工智能训练师正式成为新职业并被纳入国家职业分类目录。

一、认识人工智能训练师

从数据标注到智能系统设计，所有与 AI 训练有关的职业都可以被称为“人工智能训练师”。人工智能训练师包含数据标注员、人工智能算法测试员两个工种，按照《人

工智能训练师国家职业技能标准》从低到高一共分为五级：五级 / 初级工、四级 / 中级工、三级 / 高级工、二级 / 技师、一级 / 高级技师。

人工智能训练师的主要任务包括以下五个方面。

- (1) 标注和加工图片、文字、语音等业务的原始数据。
- (2) 分析提炼专业领域特征，训练和评测人工智能产品相关算法、功能和性能。
- (3) 设计人工智能产品的交互流程和应用解决方案。
- (4) 监控、分析、管理人工智能产品应用数据。
- (5) 调整、优化人工智能产品参数和配置。

简而言之，人工智能训练师的工作就是让 AI 更“懂”人类，他们根据客户的需求，将数据“喂”给机器人，并不断调教优化机器人，使其“通情理、懂人性”，从而更好地为人类服务。我们熟悉的天猫精灵、百度小度、智能客服等智能产品背后都有人工智能训练师的功劳。随着人工智能时代的到来，各行各业都在进行智能化升级，对人工智能训练师的需求日益增长，人工智能训练师的规模将迎来爆发式的增长。

二、人工智能训练师典型工作任务

（一）数据标注

数据标注是使用工具对未经处理的图片、文本、语音及其他原始数据进行加工处理，并进一步转换为机器可识别信息的过程。加州科技大学校长曾表示，机器识别事物主要是通过物体的一些特征进行识别，而被识别的物体还需要通过数据标注才能让机器知道这个物体是什么。简单来说，数据标注就是数据标注人员借助标记工具对原始数据进行加工的一种行为。

数据标注分为图像类、文本类、语音类及其他类四种基本类型。图像类数据标注主要包括矩形拉框、多边形拉框、关键点标注等标注形式；文本类数据标注主要包括情感标注、实体标注等标注形式；语音类数据标注包括语音转写等标注形式；其他类数据标注则形式比较灵活与多变。

当下的人工智能也被称为数据智能。因为要想实现人工智能，需要投入极大量的训练数据，因此，数据也被称作人工智能的血液。但是对于深度学习来讲，仅有数据是远远不够的，数据只有加上标签，才能用于机器的训练、学习和进化。所以，数据标注工作就显得十分重要，这也是数据标注员被称作“人工智能背后的人工”的原因。

（二）模型训练

人工智能的模型本质上就是训练机器用不同的算法来掌握一个个不同的规则，然后举一反三。我们可以把机器想象成一个小孩子，人工智能训练师通过“喂”他猫和狗的图像数据教他认识哪个动物是狗，哪个动物是猫，久而久之，小孩子就能产生认知模式，有了这个认知模式他就能辨认出猫和狗了。这个学习过程就叫模型训练，训练产生的认知模式就是猫狗分类模型。小孩子有了猫狗分类模型后，跑过来一只动物，我们问他“这是狗吗”，他会回答“是”或者“否”，这个过程就是模型的应用——预测。



人工智能训练师证书的申报条件

不同等级的人工智能训练师证书的申报条件详见表 1-1。

表 1-1 人工智能训练师证书的申报条件

职业技能等级	申报条件
五级 / 初级工	(1) 累计从事本职业或相关职业工作 1 年（含）以上。 (2) 本职业或相关职业学徒期满
四级 / 中级工	(1) 取得本职业或相关职业五级 / 初级工职业资格证书（技能等级证书）后，累计从事本职业或相关职业工作 3 年（含）以上。 (2) 累计从事本职业或相关职业工作 4 年（含）以上。 (3) 取得技工学校本专业或相关专业毕业证书。 （含尚未取得毕业证书的在校应届毕业生，或取得经评估论证、以中级技能为培养目标的中等及以上职业学校本专业或相关专业毕业证书）
三级 / 高级工	(1) 取得本职业或相关职业四级 / 中级工职业资格证书后，累计从事本职业或相关职业工作 3 年（含）以上。 (2) 取得本职业或相关职业四级 / 中级工职业资格证书，并具有高级技工学校、技师学院毕业证书（含尚未取得毕业证书的在校应届毕业生）；或取得本职业或相关职业四级 / 中级工职业资格证书，并具有经评估论证、以高级技能为培养目标的高等职业学校本专业或相关专业毕业证书（含尚未取得毕业证书的在校应届毕业生）。 (3) 具有大专及以上学历本专业或相关专业毕业证书，累计从事本职业或相关职业工作 3 年（含）以上
二级 / 技师	(1) 取得本职业或相关职业三级 / 高级工职业资格证书后，累计从事本职业或相关职业工作 4 年（含）以上。 (2) 取得本职业或相关职业三级 / 高级工职业资格证书的高级技工学校、技师学院毕业生，累计从事本职业或相关职业工作 3 年（含）以上；或取得本职业或相关职业预备技师证书的技师学院毕业生，累计从事本职业或相关职业工作 2 年（含）以上
一级 / 高级技师	取得本职业或相关职业二级 / 技师职业资格证书后，累计从事本职业或相关职业工作 4 年（含）以上

思考与练习

一、填空题

1. 人工智能训练师包含 _____、_____ 两个工种，从低到高一共分为五级：_____、_____、_____、_____、_____。

2. 我们熟悉的天猫精灵、百度小度、智能客服等智能产品背后都有 _____ 的功劳。

3. 数据标注是使用工具对 _____ 及其他 _____ 进行加工处理。

4. 人工智能训练师工作流程包括 _____、_____、_____、_____、_____。

二、判断题

1. 人工智能训练师的工作就是让 AI 更“懂”人类。()

2. 数据标注分为四种基本类型，包括图像类、文本类、语音类及其他类。()

3. 人工智能算法测试员被称作“人工智能背后的人工”。()

4. 人工智能的模型本质上就是训练机器用不同的算法来掌握一个个不同的规则。()

三、简答题

人工智能训练师的主要任务包括哪几个方面？

任务四

人工智能应用支撑认知

任务描述

无论是各种智能穿戴设备，还是各种进入家庭的陪护、安防、学习机器人、智能家居、医疗系统，这些改变我们生活方式的新事物都是人工智能的研究与应用成果。是什么样的技术来支撑这广泛且庞大应用领域的呢？越深入了解，李强的兴趣越浓，本任务他将带我们一起去学习人工智能技术的三大基础支撑——算力、算法、数据，如图 1-42 所示。

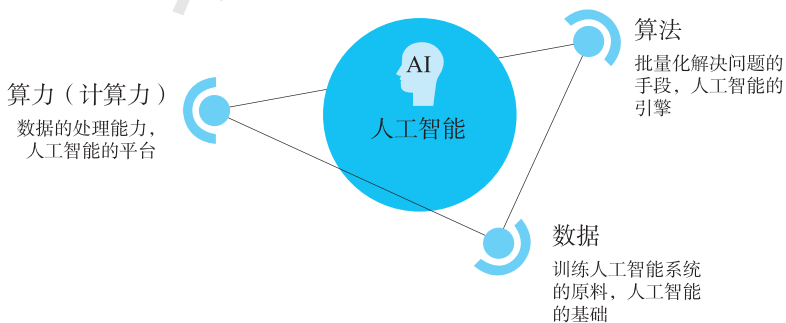


图 1-42 人工智能的三大基础支撑

任务目标

1. 能根据学习和工作获取算力。
2. 能分析实际应用场景，选择合适类型的算法模型。
3. 会使用工具进行数据标注。

任务实施

一、人工智能技术基础支撑——算力

2016年，世界顶级围棋高手李世石与AI进行围棋对决，最后竟以1:4惨败于谷歌AlphaGO。人工智能凭什么能够战胜人类？是超级计算机的算力为AI提供了支撑。当时的AlphaGo消耗了176个GPU和1202个CPU的硬件设备计算资源。在人工智能技术飞速发展的今天，功能强大的计算机显得不可或缺。AI通过计算机强大的算力处理大量的相关数据，并以神经网络不断学习成长，最终获得技能，战胜人类选手，算力对于AI的重要性不言而喻。中国工程院院士、浪潮集团首席科学家王恩东认为，人类社会已经快速步入智慧时代，计算力是这个时代的核心驱动力、生产力。

(一) 人工智能的算力

算力（也称作计算力），是设备的计算能力，也是数据处理的能力。AI的许多数据处理涉及矩阵乘法和加法。不管是图像识别等领域常用的算法，还是语音识别、自然语言处理等领域的算法，本质上都是矩阵或vector的乘法、加法，然后配合一些除法、指数等算法。CPU可以拿来执行AI算法，但因为其内部有大量其他逻辑，而这些逻辑对于目前的AI算法来说是完全用不上的，所以CPU并不能达到最优的性价比。因此，具有海量并行计算能力、能够加速AI计算的AI芯片应运而生。以4GHz、128bit的POWER8的CPU为例，假设要处理的是16bit的数据，该CPU理论上每秒可以完成 $16 \times 4G = 64$ 次。以国际巨头公司谷歌的AI芯片TPU1为例，主频为700MHz，有 $256 \times 256 = 64k$ 个乘加单元，每个时间单元可同时执行一个乘法和一个加法，那就是128k个操作。可以看出，在AI算法处理上，AI芯片比CPU快大约1000倍。如果训练一个模型的话，TPU需要花费1个小时，但放在CPU上则要41天。综上所述，人工智能的算力主要取决于芯片。

(二) AI芯片

从广义范畴上讲，面向AI计算应用的芯片都可以称为AI芯片。在狭义上，AI芯片指专门针对AI算法做了特殊加速设计的芯片，以GPU、FPGA、ASIC为代表的AI芯片，是基于传统芯片架构对某类特定算法或者场景进行AI计算加速的芯片，这三种芯片也是目前可大规模商用的主流AI芯片。这三种主流AI芯片分别对应的供应商如图1-43所示。

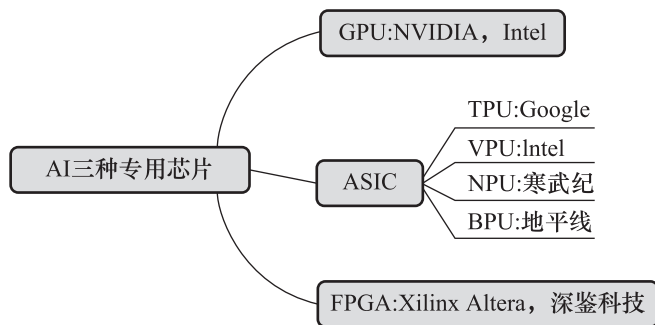


图 1-43 AI 三种专用芯片对应的供应商

1. GPU

GPU (Graphics Processing Unit)，即图形处理器，采用由大量核心组成的大规模并行计算架构，专为同时处理多重任务而设计。由于其具有良好的矩阵计算能力和并行计算优势，最早被用于 AI 计算，在数据中心中获得了大量应用。GPU 采用并行架构，超过 80% 部分为运算单元，具备较高的运算速度。相比之下，CPU 仅有 20% 为运算单元，更多的是逻辑单元，CPU 擅长逻辑控制与串行运算，而 GPU 擅长大规模并行运算，但是 GPU 无法单独工作，必须由 CPU 进行控制调用。GPU 目前是最主流的通用性 AI 芯片。在通用性芯片领域，除了英特尔和 AMD 的 CPU 外，美国的英伟达公司 (NVIDIA) 是行业龙头，几乎垄断了人工智能的 GPU 市场。

2. FPGA

FPGA (Field-Programmable Gate Array)，即现场可编程门阵列，作为专用集成电路领域中的一种半定制电路出现，适用于多指令、单数据流的分析，常用于推理阶段。FPGA 灵活性较好，处理简单指令、重复计算的能力比较强，此外，在云计算架构形成 CPU+FPGA 的混合异构中，比 GPU 具有更低的损耗和更高的性能，适用于高密度计算，在深度学习的推理阶段有着更高的效率和更低的成本。国外包括亚马逊、微软都推出了基于 FPGA 的云计算服务，国内包括腾讯云、阿里云早在 2017 年便推出了基于 FPGA 的服务，百度大脑也使用了 FPGA 芯片，而被 Xilinx 收购的深鉴科技也是基于 FPGA 来设计深度学习的加速器架构的，并将其灵活扩展，用于服务器端和嵌入式端。

3. ASIC

ASIC (Application Specific Integrated Circuits)，即专用集成电路，是一种为专用目的设计的、面向特定用户需求的定制芯片，图 1-44 中的芯片就是我国寒武纪公司生产的 ASIC 芯片。GPU 和 FPGA 除了是一种技术路线之外，还是实实在在的确定的产品。与 GPU 和 FPGA 不同，ASIC 就是一种技术路线或者方案，其呈现出的最终形态与功能也是多种多样的。近年来涌现出的类似 TPU、NPU、VPU、BPU 等令人眼花缭乱的各种芯片，本质上都属于 ASIC。ASIC 不具备 GPU 和 FPGA 的灵活性，定制化的 ASIC 一旦制造完成将不能更改，所以其初期成本高，开发周期长，因此，其进入门槛也高。目前，ASIC 大多由具备 AI 算法又擅长芯片研发的巨头参与，如 Google 研发出了 TPU。

由于完美适用于神经网络相关算法，ASIC 在性能和功耗上都要优于 GPU 和 FPGA，TPU1 的性能是传统 GPU 的 14~16 倍，NPU 的性能则是 GPU 的 118 倍。

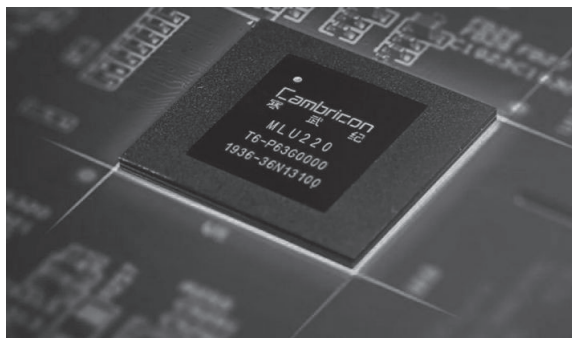


图 1-44 寒武纪公司生产的 ASIC 芯片

除了以上已经达到商用规模的 AI 芯片，还有类脑芯片、可重构通用 AI 芯片等较为前沿的芯片。类脑芯片在架构上类似人类大脑，用于进行神经拟态计算，完全开辟了另一条实现人工智能的道路，而不是作为人工神经网络或深度学习的加速器存在。目前该类芯片的状态还只是小规模的研究与应用，代表产品有 IBM 的 TrueNorth 和清华大学的“芯机”系列芯片。可重构通用 AI 芯片遵循软件定义芯片思想，可重构计算技术，允许硬件架构和功能随软件变化而变化，兼具处理器的通用性和 ASIC 的高性能和低功耗，是实现软件定义芯片思想的核心，被公认为突破性的下一代集成电路技术。清华大学微电子学研究所设计的 AI 芯片 Thinker，采用可重构计算架构，支持卷积神经网络、全连接神经网络和递归神经网络等多种 AI 算法。

超级计算机是目前世界上功能最强大的计算机。与普通的个人计算机不同，超级计算机的最终竞争指标之一是计算能力。根据 2022 年最新统计，排名前两位的超级计算机都来自中国，排名第一的是无锡国家超级计算中心的“神威太湖之光”，排名第二的是广州国家超级计算机中心的“天河二号”，排名第三至第五的超级计算机都来自美国，十大超级计算机中有一半来自美国。

（三）AI 芯片应用

AI 芯片部署的位置有云端、终端两种。云端，即数据中心，在深度学习的训练阶段需要极大的数据量和运算量，单一处理器无法独立完成，因此，训练环节只能在云端实现。终端，即手机、安防摄像头、汽车、智能家居设备、各种 IoT 设备等执行边缘计算的智能设备。根据部署位置，可以分为云 AI 芯片和端 AI 芯片。

AI 的实现包括训练、推理两个环节。训练，是指通过大数据训练出一个复杂的神经网络模型，即用大量标记过的数据来“训练”相应的系统，使之可以适应特定的功能。训练需要极高的计算性能，需要较高的精度，需要能处理海量的数据，需要有一定的通用性，以便完成各种各样的学习任务。推理，是指利用训练好的模型，使用新数据推理出各种结论，即借助现有神经网络模型进行运算，利用新的输入数据来一次性获得正确结论的过程，也称为预测或推断。所以根据承担任务的不同，AI 芯片可以分为用于构建

神经网络模型的训练芯片和利用神经网络模型进行推理预测的推理芯片。

云端与终端所使用的对应的芯片分类如表 1-2 所示。

表 1-2 云端与终端所使用的对应的芯片分类

	训练	推理
云端	GPU: NVIDIA, AMD FPGA: Intel, Xilinx ASIC: Google	GPU: NVIDIA FPGA: Intel, Xilinx, 亚马逊, 微软, 百度, 阿里, 腾讯 ASIC: Google, 寒武纪, 比特大陆, WaveComputing, Groq
终端	—	ASIC: 寒武纪, 地平线, 华为海思, 高通, ARM FPGA: Xilinx GPU: NVIDIA, ARM

(四) 智算中心

随着数据总量的增长和智能化社会构建需求的扩大,人工智能产业对算力的要求越来越高。中国工程院院士、浪潮集团首席科学家王恩东认为,在新基建各大领域之中,相比云计算和大数据,人工智能对算力的需求几乎是“无止境”的。根据人工智能研究组织 Open AI 统计,从 2012 年至 2019 年,随着人工智能深度学习“大深多”模型的演进,模型计算所需的计算量已增长了 30 万倍。斯坦福大学发布的《AI Index 2019》报告也显示,2012 年以后,算力需求每三四个月就会翻一番,经常会顾此失彼。随着新基建的加速建设,人工智能与大数据、云计算、物联网等的融合也会进一步加快,智慧医疗、无人驾驶、智慧城市、智慧金融等应用场景,背后都需要算力支撑。浪潮集团人工智能和高性能计算部总经理刘军认为,如果算力不能快速提升,那我们将不得不面临一个糟糕的局面:当规模庞大的数据用于人工智能的训练学习时,数据量将超出内存和处理器的承载上限,整个训练过程将变得无比漫长,甚至完全无法实现最基本的人工智能。

近年来,已有不少超算中心运用了人工智能芯片和服务器来强化其算力,用于提升对人工智能产业的服务能力,也可以说是对传统超算中心进行“AI 化”。2019 年,西安的沣东新城搭建了西北地区首个人工智能领域的新型基础设施——“沣东人工智能计算创新中心”;由中科院、广东省、珠海市、横琴新区共同建设的“横琴先进智能计算平台”,是粤港澳大湾区首个先进智能计算平台,被列入广东省政府 2019 年工作报告,并纳入了广东省委、省政府印发的《关于贯彻落实〈粤港澳大湾区发展规划纲要〉的实施意见》。

2020 年 4 月 20 日,国家发展改革委首次明确了新型基础设施的范围。新型基础设施主要包括信息基础设施、融合基础设施、创新基础设施三个方面。其中:信息基础设施主要是指基于新一代信息技术演化生成的基础设施,比如,以 5G、物联网、工业互联网、卫星互联网为代表的通信网络基础设施;以人工智能、云计算、区块链等为代表的新技术基础设施;以数据中心、智能计算中心为代表的算力基础设施等。智能计算中心被明确纳入了新基建的范围。

2020年11月17日,国家信息中心信息化和产业发展部联合浪潮发布了《智能计算中心规划建设指南》,其中对智能计算中心的规划建设给出了清晰的指导,并对智能计算中心进行了明确定义:基于最新人工智能理论,采用领先的人工智能计算架构,提供人工智能应用所需算力服务、数据服务和算法服务的公共算力新型基础设施,通过算力的生产、聚合、调度和释放,高效支撑数据开放共享、智能生态建设、产业创新聚集,有力促进AI产业化、产业AI化及政府治理智能化。

二、人工智能技术基础支撑——算法

做饭机器人一直是人类的追求之一,如何让机器人掌握不同菜系的烹饪方法就是实现这一追求的算法。我们通过量化各个步骤,分析各种情况,给出不同反应,指导机器人学习人类的烹饪过程就是复杂的算法设计过程。

(一) 算法的定义

算法(Algorithm)是指在解决某个问题的时候,按照某种计算方法及步骤进行处理的过程。计算机的算法则是指计算机解决某个问题以及按照何种方法进行判断的方案,还包括计算的一系列指令。

(二) 机器学习的概念

机器学习(Machine Learning, ML)是一门多领域交叉学科,涉及统计学、概率论、计算机科学等多门学科。机器学习用于研究怎样让计算机具备像人类一样的学习能力,通过数据或经验不断优化计算机算法的性能,是人工智能的核心部分。人类的学习是一个人根据过往的经验,对一类问题形成某种认识或总结出一定的规律,然后利用这些知识来对新的问题下判断的过程。因此,机器学习是指用某些算法指导计算机利用已知数据学习得出适当的模型,并应用此模型对新的情况给出判断。

机器学习根据学习方式的不同可以分为监督学习、无监督学习、强化学习三类。监督学习与其他两类学习不同之处在于其输入的训练样本带有正确输出的标记,其他两类的样本则没有输出标记,强化学习是计算机与环境的交互过程中以提高算法性能为标准不断优化算法的模型。

1. 监督学习

监督学习又称有导师的学习,输入的训练样本带有输出标记,计算机根据输入和输出结果之间的关系,不断调整模型使其输出与样本标记一致。

2. 无监督学习

无监督学习是指输入的训练样本没有输出标记,计算机需要从样本中找出数据的内置结构,抽取出通用的规则。无监督学习通常通过聚类的方式来处理复杂的数据。

3. 强化学习

强化学习是指计算机在与环境的交互过程中,通过统计环境的反馈,动态规划模型来达到最优性能的一类学习方法。强化学习通过试错的方式,根据不同的选择得到的反馈来判断性能,不断调整解决方案。

机器学习发展到今天，产生了很多优秀的算法，图 1-45 列出了按照算法的学习方式分类的几种经典的机器学习算法，以便人们在建模和对算法选择的时候可以根据输入数据选择最合适的算法来获取最好的结果。

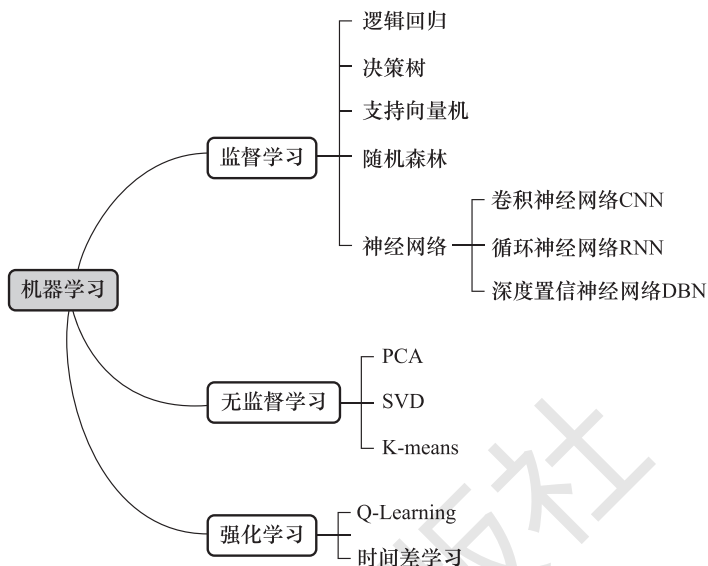


图 1-45 常见机器学习算法分类

（三）深度学习的概念

深度学习（Deep Learning，DL）是机器学习的一个子集，是指通过模仿人类大脑的思考和学习方式，学习样本数据的内在规律和表示层次，最终让机器能够像人一样具有分析学习能力，能够识别文字、图像和声音等数据的学习方法。人工智能、机器学习、深度学习三者间的关系如图 1-46 所示。

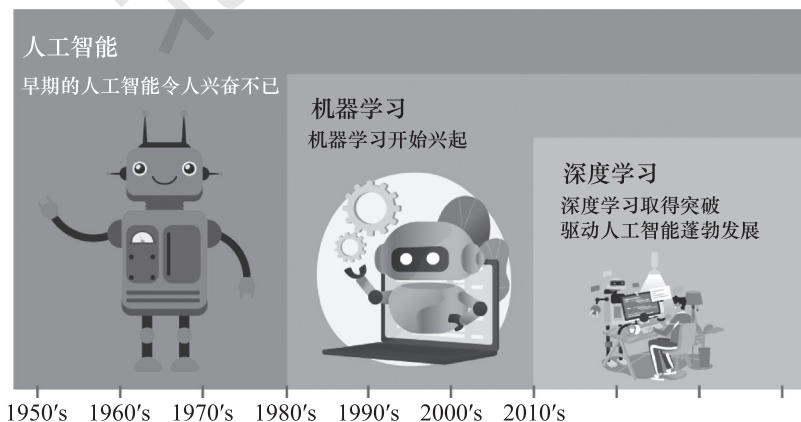


图 1-46 人工智能、机器学习、深度学习三者间的关系

（四）神经网络的概念

生物神经网络是指由生物的大脑神经元、细胞、触点等组成的网络。深度学习通

过研究模仿生物的大脑的神经网络，从而建立一个类似的学习策略，即人工神经网络（Artificial Neural Networks, ANNs），它是模拟生物神经系统建立的计算机模型。模仿人类神经元设计的神经元模型如图 1-47 所示。

由图 1-47 可以看到生物大脑的基本单元——神经元的组成，它以细胞体为主体，由许多向周围延伸的不规则树枝状纤维进行连接，构成神经网络。模仿生物神经元设计的神经元模型是一个包含输入、输出与计算功能的模型。输入可以类比为神经元的树突，而输出可以类比为神经元的轴突，计算则可以类比为细胞核。大量的人工神经元（节点）连接构成神经网络，如图 1-48 所示。

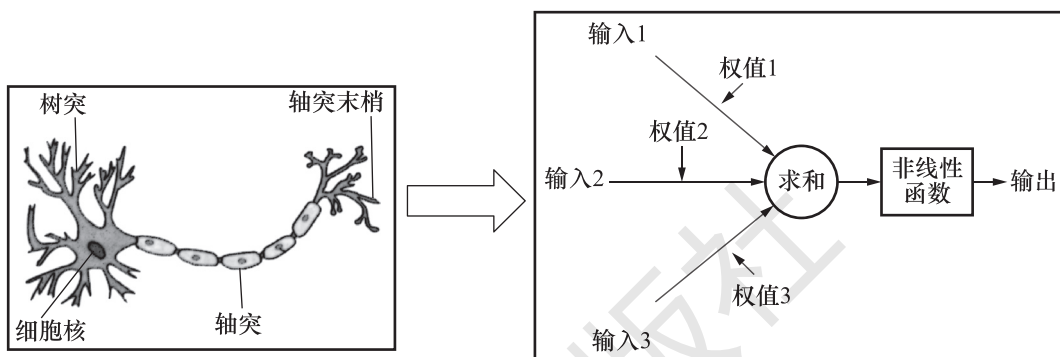


图 1-47 模仿人类神经元设计的神经元模型

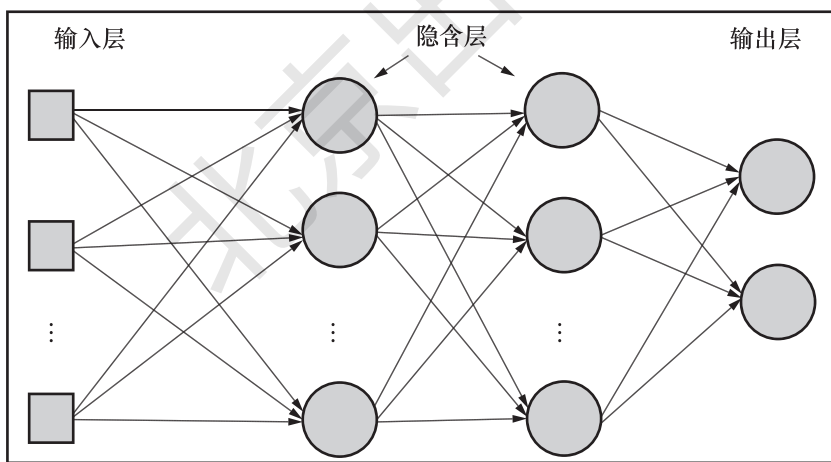


图 1-48 神经网络结构图

神经网络是由层构成的，一般包括输入层、输出层、隐含层。每个节点的输出函数为 $y = \alpha(W \cdot x + b)$ ，其中 x 是输入向量， y 是输出向量， $\alpha()$ 是激活函数， b 是偏移向量， W 是权值矩阵，每一层将输入 x 通过以上公式得到输出 y 。激活函数给神经元引入了非线性因素，使得神经网络可以任意逼近任何非线性函数。每个神经元之间的连线对应一个权值，神经网络中的所有权值组成一个权值矩阵 W 。

神经网络在外界输入信息的基础上改变了内部结构，对输入和输出之间复杂的关系

进行建模，自适应地探索数据的模式。神经网络的学习过程为：把训练集中的每个输入加到神经网络中，神经网络根据预测值与目标值之间的误差（通过损失函数计算得到）不断调整网络各层的权值矩阵（ W ）与偏移向量 b ，使神经网络的预测值与目标值一致，待各层权值都收敛到一定值时，学习过程结束。然后我们就可以用生成的神经网络来对未知数据进行判断。

神经网络发展到今天已有感知器、Hopfield 网络、玻尔兹曼机、反向传播网络等一百多种网络模型。根据网络连接的结构，神经网络模型可以分为前向神经网络和反馈神经网络。前向神经网络中各个神经元接收前一级的输入，并输出到下一级，网络中没有反馈，可以用一个有向无环路图表示。反馈神经网络（BP 网络）是指网络内的神经元之间有反馈，可以根据误差逆向调整参数的神经网络。

2010 年，深度学习得到了迅猛的发展，典型的神经网络模型有卷积神经网络（Convolutional Neural Networks, CNN）、循环神经网络（Recurrent Neural Network, RNN）、长短时记忆神经网络（Long/Short Term Memory Network, LSTM）、深度信念神经网络（Deep Belief Network, DBN）等。

深度神经网络的训练过程主要包括以下五步。

第一步：输入样本到网络，初始化参数，对权值矩阵 W 、偏移向量 b 进行随机初始化；

第二步：进行前向传播，得到输出；

第三步：通过损失函数计算实际输出与理想输出之间的差值；

第四步：反向传播，以极小化误差为标准调整权值矩阵；

第五步：梯度下降，更新参数。

在神经网络的训练中，有三个重要的概念：Epoch、Batch Size 和迭代。

（1）Epoch：当一个完整的数据集通过了神经网络一次并且返回了一次，这个过程称为一个 Epoch。

（2）Batch Size：当一次输入一个的数据集对于计算机而言太庞大的时候，就需要把它分成多个小块，也就是将数据集分成几个 Batch。Batch Size 则是一个 Batch 中的样本总数。

（3）迭代：Batch 需要完成一个 Epoch 的次数。

Tensorflow 是目前流行的神经网络搭建平台，被广泛用于各类机器学习算法的编程实现。Tensorflow Playground 展示了数据是如何“流动”的，能让你在浏览器中运行真实的神经网络，并且可以点击按钮和调整参数，从而了解神经网络是怎么工作的。

三、人工智能技术基础支撑——数据

实现一个功能完善的人工智能应用的过程中，数据是一个重要因素。就像要做一道好菜必须要有好的原材料一样，数据就相当于人工智能应用的“原材料”。开发一个卓越的人工智能应用就要获取合适的原材料（即数据）并对其进行加工。

（一）数据采集

如果我们做菜的时候只有白菜、豆腐之类的原材料，那么无论如何也是做不出红烧肉的。同样，在数据处理流程中，采集的数据决定了数据分析挖掘的上限。

比如，如果想了解 App 的终端用户在使用什么样的手机，就需要去采集终端用户所用设备的信息，如设备的品牌、型号等。当我们拿到这些信息数据以后，才能够去分析终端用户所使用的品牌分布和设备分布，从而了解用户设备的使用情况。

（二）数据的采集源及采集方式

在当今的大数据时代，数据的采集源往往是高度多样化的，而不同的数据源也往往需要不同的采集手段来对其进行针对性的采集。一般来讲，数据源包括但不限于如下几类。

第一类是终端数据。终端数据即在客户端或者服务器端产生的数据，如用户点了哪些页面或内容。这类数据一般采用埋点的方法进行采集。埋点是指针对用户特定的行为进行跟踪与捕捉的过程，这些被捕捉到的行为经统计后常常会用于后续对产品的迭代与优化。

第二类是开放数据。开放数据指的是开放给所有人的数据，如网页的内容数据，或者特定行业的公开数据。这类数据一般使用爬虫技术来采集。

第三类是专有平台的数据。如微信公众号的数据就存放在微信平台。我们可以通过微信提供的规范 API 接口服务来调取这个公众号自身的数据。

第四类是物理数据。物理数据是指用户在物理世界中所产生的数据。这类数据的采集往往要通过传感器来进行 AIDC 采集。AIDC 采集的全称为自动识别和数据捕获，是一种自动识别和收集数据对象，然后将其存储在计算机中的方法。如射频识别、条形码磁条扫描、GPS 传感器等都属于用于识别与捕获物理数据的方法。

第五类是主观性数据。比如，通过用户调研或是访谈的方式，收集用户的态度或是意愿。这是一种传统数据的采集方式。

第六类是数据库的数据。比如，对于一些现成的知识库，自己建设的话可能费时费力，我们可以直接通过购买的方式来拿到相应的知识库的数据。

（三）采集的数据类型划分

数据采集之后，我们需要对采集的数据进行分类整理。通常情况下，数据可以被分为非结构化数据、结构化数据以及半结构化数据三种类型。

非结构化数据指的是无法定义结构的数据。常见的非结构化数据为文本信息、图像信息、视频信息以及声音信息等等，它们的结构都千变万化，不能用一个二维表来描述。

结构化数据往往被称为行数据，是由二维表结构来逻辑表达和实现的数据，其严格地遵循数据格式与长度规范，主要通过关系型数据库进行存储和管理。如在大学生的选课系统中，学生、课程、选课、导师等数据都可以抽象为结构化数据。

半结构化数据，是指结构变化很大的结构化数据。同一类结构化数据的不同实体数

据的结构可能会有一定程度的不同，即不同实体所具有的属性会有一定程度的不同。对于这些实体来说，不同的属性之间的顺序并不重要。一个经典的半结构化数据的例子即简历信息。每一份简历都遵循着简历这个大类所存在的物理意义，即展示我们迄今为止在所在领域的成就，所以我们的简历中很有可能会有教育背景、工作经验、姓名以及联系方式等等。然而在这个大前提下，每一份简历所具有的属性都不尽相同，有的人会在简历中加入志愿者经历，有的人会加入自己所掌握的技能，有的人会加入自己的获奖经历等等，这就是我们刚刚所说的数据的结构变化很大的一个体现。半结构化数据往往以 XML 或者 JSON 等方式出现。

（四）数据预处理

在进行数据挖掘时，海量的原始数据中存在大量有缺失、异常甚至是不一致的数据，严重影响了建模的执行效率以及正确性。数据预处理的主要内容包括数据清洗、数据集成、数据变换、数据规约。预处理一方面可提高数据的质量，另一方面可使数据更好地适应特定的挖掘技术或工具。

1. 数据清洗

数据清洗主要是删除原始数据集中的无关数据、重复数据，平滑噪声数据，筛选掉与挖掘主题无关的数据并且处理缺失值、异常值等。

一般处理缺失值的时候，通常采用删除记录、数据插补，或者不处理的方式。如果删除少量数据就可以达到目标，当然最好，但是它是以减少历史数据来换取数据的完备，会造成大量资源浪费，丢弃大量隐藏在记录里的信息，所以不太推荐这种方式，而更推荐用插补的方式来补齐数据。插补的方法有很多，例如，通过均值等属性数据填补，也可以使用固定的值、临近的值进行插补。

2. 数据集成

数据挖掘之前，数据集往往是在不同的数据源中的，这时候需要将多个数据源合并存储到一个数据仓库中。由于多个数据源的表达形式是不一样的，有可能不匹配，因此，要考虑到识别的问题以及属性冗余的问题，从而将源数据在最底层加以转换、提炼和集成。这个步骤主要做的就是实体的识别以及冗余属性的识别。

实体识别是从不同数据源识别出现实世界的实体，需要做的是统一不同数据的矛盾之处。简单来说，就是将不同的字段名字统一，以及将数据的计量单位统一。例如：A 在一个数据源中叫 gender，在另一个中叫 sex，需要将 A 的名字进行统一；有的计量单位用的是 m，有的用的是 km，计量单位也需要统一。

冗余属性的识别是为了解决同一属性多次出现、同一属性命名不一致所导致的重复问题。

3. 数据变换

数据变换主要是对数据进行规范化的处理，以适应挖掘任务和算法的需要。数据变换是数据挖掘中至关重要的一个步骤。数据变换主要包括以下几个方面。

（1）简单的数据变换。

一般来说，简单的函数变换是对原始数据进行简单变换，基本使用平方、开平方等运算。简单的函数变换常用来将不具备正态分布的数据变换成具有正态分布的数据。

（2）规范化。

数据规范化处理是数据挖掘的一项基础工作，主要是消除指标之间的量纲和取值范围的差异影响，需要进行标准化处理，将数据按照比例进行缩放，从而使其落入一个特定的区域，以便于分析。

（3）连续属性离散化。

一些数据挖掘算法，特别是分类算法，要求数据是分类属性形式。如此一来，在使用这类分类算法时，需要将一些连续属性变换成分类属性，也就是连续属性离散化。这其实就是在数值的取值范围内设定若干个离散的划分点，将取值范围划分为一些离散化的区间，最后用不同的符号或数值代表每个区间的数，即需要确定分类数以及将连续属性值映射到这些分类值中。

（4）属性构造。

在数据挖掘过程中，为了提取更有用的信息，挖掘更深层次的模式，提高挖掘结果的精度，我们需要利用已有的属性来构造新的属性，并加入现有属性的集合中。下面进行一个简单的举例：通过一个餐馆的日营业总额数据，还有每日单据数量，我们就可以知道人均每笔消费的数据，从而加入新的一列中进行统计。这个数据虽然是从已有数据中延伸来的，但是直接生成新属性统计时更直观。

4. 数据规约

数据规约本质上就是在不损害数据完整性的前提下缩小数据集，使之能够高效快速地挖掘出结果。数据规约一般要从两个方向进行，一个是属性规约，另一个是数值规约。

属性规约通过属性合并来创造新属性维数，或者通过删除不相干属性来减少数据的维数，从而提高数据挖掘的效率以及降低计算成本，其目标是寻找出最小的属性子集，并确保新数据子集的概率分布尽可能接近原数据的概率分布。

数值规约指通过选择替代的、较小的数据来减少数据量，包括有参数和无参数两类。有参数方法是使用一个模型来评估数据，只存放参数，而不需要存放实际数据。无参数方法就需要存放实际数据，如直方图、聚类、抽样等。例如，聚类分析的方式就是将数据分为簇，让一个簇中的对象相互相似，反之相异，并用数据簇来替换原始数据的数据规约方式。

（五）数据标注

训练机器学习和深度学习模型，需要丰富的数据，以便将其用于部署、训练和调整模型。训练机器学习和深度学习模型需要大量经过仔细标注的数据。标注原始数据并准备将其应用于机器学习模型和其他 AI 工作流的过程称为数据标注。

要使 AI 模型做出决策并采取行动，就必须对其进行训练以理解特定的信息。训练

数据必须针对特定用例，并对其予以适当分类和标注。有了高质量的人工标注数据，我们就可以构建和改进 AI 应用，如产品推荐、相关搜索引擎结果、计算机视觉、语音识别、聊天机器人等。

1. 数据标注的分类

数据标注的分类如表 1-3 所示。

表 1-3 数据标注的分类

分类方式	分类方法	概念	优点	缺点
标注对象	图像标注	图像标注和视频标注统称为图像标注	使人脸识别和智能驾驶等技术得到发展和完善	相对复杂，耗时
	语音标注	需要人工将语音内容转录为文本内容，然后通过算法模型识别转录后的文本内容	帮助人工智能领域中的语音识别功能更加完善	算法无法直接理解语音内容，需要进行文本转录
	文本标注	与语音标注有些相似，都需要通过人工识别转录成文本形式	减少了文本识别行业和领域的人工工作量	人工识别过程繁杂
标注的构成形式	结构化标注	数据标签必须在规定的标签候选集合内，标注者通过将标注对象与标签候选集合进行匹配，选出最合理的标签值作为标注结果	标签候选集将标注类别描述得很清晰，便于标注者选择； 标签是结构化的，利于存储和后期的统计查找	遇到具有二义性标签时往往会影响最终的标注结果
	非结构化标注	标注者在规定约束内，自由组织关键字对标注对象进行描述	给标注者足够的自由，可以清楚地表达自己的观点	给数据存储和使用带来困难，不利于统计分析
	半结构化标注	标签值采用结构化标注，而标签域采用非结构化标注	标注灵活性强，便于统计查找	对标注者的要求高，工作量高，耗时
标注者类型	人工标注	雇用经过培训的标注员进行标注	标注质量高	标注成本高，时间长，效率低
	机器标注	标注者通常是智能算法	标注速度快，成本相对较低	算法对涉及高层语义的对象识别和提取效果不好

目前数据标注有三种常用的划分方式：

- (1) 根据标注对象，可分为图像标注、语音标注和文本标注。
- (2) 根据标注的构成形式，可分为结构化标注、非结构化标注和半结构化标注。
- (3) 根据标注者类型，可分为人工标注和机器标注。

图像标注包括图像标注和视频标注，因为视频也是由连续播放的图像所组成的。图像标注一般要求标注人员使用不同的颜色来对不同的目标标记物进行轮廓识别，然后给相应的轮廓打上标签，用标签来概述轮廓内的内容，以便让算法模型能够识别图像中的不同标记物。图像标注常用于人脸识别、智能驾驶车辆识别等应用。

语音标注是通过算法模型识别转录后的文本内容并与对应的音频进行逻辑关联。语音标注的应用场景包括自然语言处理、实时翻译等，语音标注的常用方法是语音转写。

文本标注是指根据一定的标准或准则对文字内容进行诸如分词、语义判断、词性标注、文本翻译、主题事件归纳等注释工作，其应用场景有名片自动识别、证照识别等。目前，常用的文本标注任务有情感标注、实体标注、词性标注及其他文本类标注。

2. 数据标注的任务

常见的数据标注任务包括分类标注、标框标注、区域标注、描点标注和其他标注等。

(1) 分类标注。

分类标注是从给定的标签集中选择合适的标签分配给被标注的对象。通常，一张图可以有多个分类/标签，如运动、读书、购物、旅行等。对于文字，又可以标注出主语、谓语、宾语，名词和动词等。此项任务适用于文本、图像、语音、视频等不同的标注对象。此处以图像的分类标注为例进行说明。如图 1-49 所示，图中显示了一张室内布局图，标注者需要对沙发、茶几、灯等不同对象加以区分和识别。

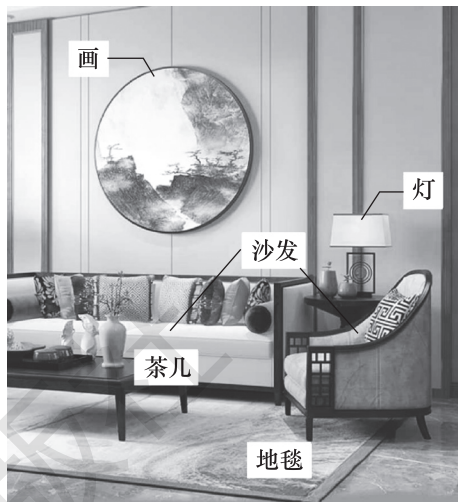


图 1-49 分类标注

(2) 标框标注。

标框标注就是从图像中选出要检测的对象，此方法仅适用于图像标注。标框标注可细分为多边形拉框和四边形拉框两种形式。多边形拉框是将被标注元素的轮廓以多边形的方式勾勒出来，不同的被标注元素有不同的轮廓，除了同样需要添加单级或多级标签以外，多边形标注还有可能会涉及物体遮挡的逻辑关系，从而实现细线条的种类识别。四边形拉框主要是用特定软件对图像中需要处理的元素（如人、车、动物等）进行一个拉框处理，同时，用一个或多个独立的标签来代表一个或多个需要处理的元素。例如，图 1-50 中对交通标识进行了多边形拉框标注，图 1-51 中对道路中的汽车进行了四边形拉框标注。



图 1-50 多边形拉框标注



图 1-51 四边形拉框标注

（3）区域标注。

与标框标注相比，区域标注的要求更加精确，而且边缘可以是柔性的，但仅限于图像标注，其主要的应用场景包括智能驾驶中的道路识别和地图识别等。在图 1-52 中，区域标注的任务是在航拍图中用曲线将城市中不同的设施的轮廓形式勾勒出来，并用不同的颜色加以区分。

（4）描点标注。

描点标注是指将需要标注的元素（比如人脸、肢体）按照需求位置进行点位标识，从而实现特定部位关键点的识别。例如，图 1-53 采用描点标注的方法对图示人物的骨骼关节进行了描点标识。人脸识别、骨骼识别等技术中的标注方法与人物骨骼关节的标注方法相同。



图 1-52 区域标注



图 1-53 描点标注

（5）其他标注。

数据标注的任务除了上述四种以外，还有很多个性化的标注任务。例如，自动摘要是指从新闻事件或者文章中提取出最关键的信息，然后用更加精练的语言写成摘要。自动摘要与分类标注类似，但两者存在一定差异。常见的分类标注有比较明确的界定，比如，在对给定图片中的人物、风景和物体进行分类标注时，标注者一般不会产生歧义；而自动摘要需要先对文章的主要观点进行标注。相对于分类标注来说，自动摘要在标注的客观性和准确性上都没有那么严格，所以自动摘要不属于分类标注。

3. 常用标注数据集和标注工具

随着人工智能、机器学习等行业对标注数据的海量需求，许多企业和研究机构纷纷推出了带标注的公开数据集。为了提高数据标注效率，一些标注工具和平台也应运而生。

（1）标注数据集。

标注数据集主要可分为图像、视频、文本和语音四类标注数据集，表 1-4 描述了这些数据集的类别、名称、用途和特性。ImageNet、COCO 和 PASCAL VOC 是三个典型的图像标注数据集，它们广泛应用于图像分类、定位和检测的研究中。由于 ImageNet 数据集拥有专门的维护团队，而且文档详细，因此，它几乎成为目前检验深度学习图像领域算法性能的“标准”数据集。COCO 数据集是在微软公司赞助下生成的数据集，除了图像的类别和位置标注信息外，该数据集还提供图像的语义文本描述，所以它也成为

评价图像语义理解算法性能的“标准”数据集。Youtube-8M 是谷歌公司从 YouTube 上采集到的超大规模的开源视频数据集，这些视频共计 800 万个，总时长为 50 万小时，包括 4800 个类别。Yelp 数据集由美国最大的点评网站提供，包括了 470 万条用户评价、15 多万条商户信息、20 万张图片 and 12 个城市信息。研究者利用 Yelp 数据集不仅能进行自然语言处理和情感分析，还可以进行图片分类和图像挖掘。LibriSpeech 数据集是目前最大的免费语音识别数据库之一，由近 1000 小时的多人朗读的清晰音频及其对应的文本组成，它是衡量当前语音识别技术最权威的开源数据集。

表 1-4 部分常用的标注数据集

类别	数据集名称	用途	大小	开放权限
图像标注数据集	ImageNet	图像分类、定位、检测	约 1TB	是
	COCO	图像识别、分割和图像语义	约 40GB	是
	PASCAL VOC	图像分类、定位、检测	约 2GB	是
	Open Image	图像分类、定位、检测	约 1.5GB	是
	Flickr30k	图片描述	30MB	是
视频标注数据集	Youtube-8 M	理解和识别视频内容	1PB	受限
	Kinetics	动作理解和识别	约 1.5TB	是
	AVA	人类动作识别		是
	UCF101	视频分类、动作识别	6.5GB	是
文本标注数据集	Yelp	文本情感分析	约 2.66GB	是
	IMDB	文本情感分析	80.2 MB	是
	Multi-Domain Sentiment	文本情感分析	52 MB	是
	Sentiment 140	文本情感分析	80 MB	是
语音标注数据集	LibriSpeech	训练声学模型	约 60GB	是
	AudioSet	声学事件检测	80MB	是
	FMA	语言识别	约 1000GB	是
	VoxCeleb	语音识别、情绪识别	150MB	是

（2）开源数据标注工具。

在选择数据标注工具时，需要考虑标注对象（如图像、视频、文本等）、标注需求（如画框、描点、分类等）和不同的数据集格式（如 COCO、PASCAL VOC、JSON 等）。常用的开源的数据标注工具及其特点详见表 1-5，其中除了 COCO UI 和 LabelMe 工具在使用时需要 MIT 许可外，其他工具均可开源使用。大部分的开源工具都可以运行在 Windows、Linux、Mac OS 系统上，仅有个别工具是针对特定操作系统开发的（如 RectLabel），而且这些开源工具大多只针对特定对象进行标注，只有少部分工具（如精灵标注助手）能够同时标注图像、视频和文本。除了表中列举的标注工具外，市场上

还有一些具有特殊功能的标注工具，如人脸数据标注工具和 3D 点云标注工具。不同标注工具的标注结果会有一些差异，但很少有研究去关注它们的标注效率和标注结果的质量。

表 1-5 常用的开源的数据标注工具及其特点

名称	简介	运行平台	标注形式	导出数据格式
LabelImg	著名的图像标注工具	Windows、Linux、Mac	矩形	XML
LabelMe	著名的图形界面标注工具，能够标注图像和视频	Windows、Linux、Mac	多边形、矩形、圆形、多段线、线段、点	VOC 和 COCO
RectLabel	图像标注	Mac	多边形、矩形、多段线、线段、点	YOLO、KITTI、COCO、CSV
VOTT	微软发布的基于 Web 方式本地部署的标注工具，能够标注图像和视频	Windows、Linux、Mac	多边形、矩形、点	TFRecord、CSV、VoTT
LabelBox	适用于大型项目的标注工具，基于 Web，能够标注图像、视频和文本	—	多边形、矩形、线、点、嵌套分类	JSON
VIA	VGG (visual geometry group) 的图像标注工具，也支持视频和音频标注	—	矩形、圆、椭圆、多边形、点和线	JSON
COCO UI	用于标注 COCO 数据集的工具，基于 Web 方式	—	矩形、多边形、点和线	COCO
Vatic	Vatic 是一个带有目标跟踪的视频标注工具，适合目标检测任务	—	—	VOC
BRAT	基于 Web 的文本标注工具，主要用于对文本的结构化标注	Lmux	—	ANN
DeepDive	处理非结构化文本的标注工具	Lmux	—	NLP
Praat	语音标注工具	Windows、Unix、Linux、Mac	—	JSON

知识链接

人工智能技术的主要技术领域有以下五个方面。

一、数理逻辑方法与专家系统

用符号描述逻辑推理过程，并由此构成专家系统，用以进行问题求解、逻辑推理和定理证明、自然语言理解、自动程序设计、智能规划、调度与指挥等工作。

二、模糊控制技术

以数学的模糊集合理论为基础，用以研究模式识别、机器视觉、智能控制等工作。

三、计算智能与进化计算及遗传算法

模拟生物进化与遗传过程，探索生物进化与遗传过程中显现出来的结构不断优化、复杂等级不断提升、种类不断多样化、对环境的变化不断适应等现象，用以探索问题求解、机器学习、模式识别、智能规划、智能控制、智能机器人、人工生命等方面的问题。

四、分布式人工智能与多智能体技术

随着计算机和网络技术的不断发展，分布式智能系统得到重视，用以研究大系统中的智能分布性、连接性、协作性、开放性、容错性、独立性。

五、神经网络

以连接主义的思路模拟大脑神经系统处理信息的方式，构建人工神经网络，用以完成问题求解、自然语言理解、自动程序设计、机器学习、模式识别、机器视觉、智能规划、智能控制等工作。

各个技术领域都能实现多种问题的研究与解决，功能有交叉，各自擅长不同的领域。



知识拓展

人工智能的三个级别

人工智能可以分为弱人工智能、强人工智能、超人工智能三个级别，目前弱人工智能应用非常广泛。

1. 弱人工智能

弱人工智能（Artificial Narrow Intelligence, ANI）是擅长于单个方面的人工智能。如能战胜围棋世界冠军的人工智能阿尔法狗，但是它只会下围棋，如果我们问它其他的问题，它就不知道怎么回答了。像阿尔法狗一样，只具有擅长某一方面能力的人工智能就是弱人工智能。

2. 强人工智能

强人工智能（Artificial General Intelligence, AGI）是一种接近人类级别的人工智能，是在各方面都能和人类比肩的人工智能，人类能干的脑力活它都能干。创造强人工智能比创造弱人工智能难得多，我们现在还做不到。强人工智能具有抽象思维，能够思考、

计划、解决问题、理解复杂理念、快速学习和从经验中学习等。强人工智能在进行这些操作时应该和人类一样得心应手。

3. 超人工智能

科学家把超人工智能 (Artificial Super Intelligence, ASI) 定义为在几乎所有领域都比最聪明的人类大脑还聪明很多的人工智能, 包括科学创新、通识和社交技能。超人工智能可以是各方面都比人类强一点, 也可以是各方面都比人类强万亿倍的人工智能。

思考与练习

一、填空题

1. 人工智能技术对智能制造行业的支持主要表现在 _____、_____、_____。
2. AI 的许多数据处理涉及 _____ 和 _____。
3. 机器学习分为三类不同的学习方式: _____、_____、_____。
4. 采集的数据类型可以分为三种: _____、_____ 以及 _____。

二、判断题

1. AI 通过计算机强大的算力处理大量的相关数据, 并最终获得技能。()
2. 采集终端用户所用设备的信息并不能知道 App 的终端用户在使用什么样的手机。()
3. 图像识别、语音识别、自然语言处理等领域的算法, 都是矩阵或 vector 的乘法、加法。()
4. 传感器和通信设备对人居环境进行监测形成数据流。()

三、实践题

上网或到图书馆等查阅相关资料, 围绕“人工智能在你学习的专业中已有哪些应用”“你最想利用人工智能来解决什么问题”“你打算如何将人工智能与专业学习相结合”等主题展开讨论, 明晰思路, 最终写成 Word 文档形式的阐述报告。