人工智能素养与实践



类目：人工智能类
书名：人工智能素养与实践
主编：张雯晖　余 毅　胡有年
出版社：首都师范大学出版社
开本：大 16 开
书号：978-7-5656-9703-6
使用层次：通用
出版时间：2026 年 8 月
定价：49.80 元
印刷方式：四色
是否有资源：有



人工智能素养与实践

责任编辑：付宏武
封面设计：旗语书装

ISBN 978-7-5656-9705-6
定价：49.80元

人工智能创新型通识教育新形态教材

人工智能素养与实践

主编◎张雯晖　余　毅　胡有年

首都师范大学出版社
CAPITAL NORMAL UNIVERSITY PRESS

人工智能创新型通识教育新形态教材



人工智能素养与实践

主　编◎张雯晖　余　毅　胡有年

首都师范大学出版社
CAPITAL NORMAL UNIVERSITY PRESS

人工智能创新型通识教育新形态教材



人工智能素养与实践

主　编◎张雯晖　余　毅　胡有年
副主编◎肖　鹏　廖　丽　刘　柳
罗　缔

首都师范大学出版社
CAPITAL NORMAL UNIVERSITY PRESS

图书在版编目（CIP）数据

人工智能素养与实践 / 张雯晖，余毅，胡有年主编. 北京 : 首都师范大学出版社，2026. 8. -- ISBN 978-7-5656-9703-6

Ⅰ. TP18

中国国家版本馆 CIP 数据核字第 20262US315 号

人工智能素养与实践

张雯晖　余　毅　胡有年　主编

责任编辑　付宏武

出版发行　首都师范大学出版社

地　　址　北京市西三环北路 105 号

邮　　编　100048

电　　话　010-68418523（总编室）　010-68982468（发行部）

网　　址　http：//cnupn. cnu. edu. cn

印　　刷　三河市祥达印刷包装有限公司

经　　销　全国新华书店

版　　次　2026 年 8 月第 1 版

印　　次　2026 年 8 月第 1 次印刷

开　　本　880 mm×1230 mm　1/16

印　　张　13

字　　数　330 千

定　　价　49. 80 元

版权所有　违者必究

如有质量问题　请与出版社联系退换

PREFACE 前言

教育部于2024年启动了教育系统人工智能大模型示范应用行动，旨在打造人工智能通识课程体系，赋能理工医农文等各类人才培养，“人工智能＋”时代来临，高校迫切需要构建一套完整的人工智能通识课程体系，为社会培养输送大量具有AI素养的优秀专业人才。

党的二十大报告提出要实施科教兴国战略，强化现代化建设人才支撑，强调要深化教育领域综合改革，加强教材建设和管理。为了响应党中央的号召，深入贯彻全国教育大会和《教育强国建设规划纲要（2024—2035年）》精神，更好构建适应社会需求的人工智能通识课程教育体系，我们经充分调研、严谨论证，编写完成《人工智能素养与实践》。

随着ChatGPT一鸣惊人地将人工智能技术应用呈现在人们的面前，人工智能技术展现出颠覆性的力量，DeepSeek的异军突起更是在全社会掀起了人工智能应用热潮。身处人工智能时代，我们已经切身感受到了人工智能对人们的思维模式和生活方式的改变，人工智能技术必将对人类的社会生产和生活产生重大而深远的影响。

与同类书相比，本书具备以下特点：

循序渐进，由浅入深

本书为人工智能的入门教材，以人工智能的整个处理流程为主线，逐一介绍了人工智能所涉及的相关技术，帮助读者形成对人工智能知识体系及其应用领域的轮廓性认识，为读者在人工智能领域的继续深造奠定基础。

说理透彻，注重实操

本书在重视基本理论的前提下，兼顾实际可操作性，注重问题解决。全书每章均设置大量实训操作题，帮助读者巩固所学知识。

技能与思政教育紧密结合

在讲解人工智能专业知识的同时，紧密结合思政教育主旋律，从专业知识角度引申拓展，引导学生树立正确的思想价值观念。

项目式教学，实操性强

全书采用项目式教学模式，将人工智能应用知识分解并融入一个个实践操作的训练项目中，增强了本书的实用性。

由于编者能力有限，书中难免存在不妥之处，恳请读者朋友提出宝贵意见，不胜感激。

编　者

2026 年 6 月

CONTENTS 目 录

1

项目一　培养具备人工智能素养的下一代：人工智能概述　/1

【项目导读】　/1
【学习指导】　/1
任务一　认识人工智能　/2
任务二　人工智能技术基础　/6
任务三　人工智能大模型　/15
任务四　人工智能现状　/19
【思考与练习】　/26

2

项目二　探索视觉智能新境界：AI 生成图像技术基本应用　/27

【项目导读】　/27
【学习指导】　/27
任务一　AI 生成图像技术　/28
任务二　讯飞星火 AI 生图　/40
任务三　文心一言 AI 生图　/58
任务四　可灵 AI 生图　/63
【思考与练习】　/70

3 项目三　动态视界：AI 生成视频技术的创新与应用　/71

【项目导读】　/71
【学习指导】　/71
任务一　AI 生成视频技术　/72
任务二　即梦生成视频　/76
任务三　可灵生成视频　/82
【思考与练习】　/102

4 项目四　文字的未来：AI 生成文本技术的探索与实践　/104

【项目导读】　/104
【学习指导】　/104
任务一　AI 文本生成技术　/105
任务二　豆包　/109
任务三　Kimi　/123
【思考与练习】　/141

5 项目五　AI 伦理与法治　/143

【项目导读】　/143
【学习指导】　/143
任务一　隐私和数据安全　/144
任务二　AI 伦理与法治原则　/150
任务三　AI 法律规范与算法监管政策　/155
【思考与练习】　/162

6

项目六　综合应用：DeepSeek 协同多模态 AI 工具实践　/164

【项目导读】　/164
【学习指导】　/164
任务一　DeepSeek 入门与能力认知　/165
任务二　DeepSeek 生成与优化文本内容　/169
任务三　DeepSeek＋图像生成：AI 生成图像　/173
任务四　DeepSeek＋视频生成：AI 生成视频　/181
【思考与练习】　/191

附录一　装备制造大类：AI 赋能智能制造与工程设计　/192

附录二　电子信息大类：AI 赋能电子工程与软件开发　/194

附录三　能源动力大类：AI 赋能智慧能源与动力系统　/196

附录四　财经商贸大类：AI 赋能商业分析与数字营销　/198

参考文献　/200

项目一

培养具备人工智能素养的下一代：人工智能概述

项目导读

在人工智能技术迅猛发展并深刻影响社会生产与生活方式的今天，培养具备人工智能素养的下一代已成为教育的重要使命。本项目以“培养具备人工智能素养的下一代”为主题，旨在帮助学生全面认识人工智能的基本概念、核心技术与发展现状。通过系统学习人工智能的定义、术语、分类与技术架构，了解机器学习、深度学习、计算机视觉、自然语言处理等关键技术原理，学生将建立起对人工智能的科学认知。

学习指导

通过本项目学习，你将：

- 理解人工智能的基本概念、发展历程和核心要素。
- 认识人工智能的技术架构与关键技术原理。
- 了解主流人工智能大模型的特点与应用场景。
- 初步具备应用人工智能技术解决实际问题的能力。

任务一 认识人工智能

一、人工智能的定义

人工智能（Artificial Intelligence，AI）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科学。

人工智能是计算机科学分支，旨在探明智能本质，研发可模拟、复刻人类智能行为反应的新型智能机器，该领域的研究对象包括机器人、语言识别、图像识别、自然语言处理和专家系统等。人工智能从诞生以来，理论和技术日益成熟，应用领域也不断扩大，可以设想，未来人工智能带来的科技产品，将会是人类智慧的“容器”。人工智能可以模拟人类意识、思维的信息处理过程。人工智能不是人的智能，但能像人那样思考，也可能超过人的智能。

人工智能在计算机上实现时有 2 种不同的方式。

一种是采用传统的编程技术，使系统呈现智能的效果，而不考虑所用方法是否与人或动物机体所用的方法相同。这种方法叫工程学方法（Engineering Approach），它已在一些领域取得了成果，如文字识别、电脑下棋等。采用这种方法，需要人工详细规定程序逻辑，人工编程就非常烦琐，容易出错。而一旦出错，就必须修改原程序，重新编译、调试，最后为用户提供新版本或者新补丁，操作繁琐。

另一种是模拟法（Modeling Approach），它不仅要看效果，还要求实现方法也和人类或生物机体所用的方法相同或相类似。编程者要为每一角色设计一个智能系统（一个模块）来进行控制。利用这种方法来实现人工智能，要求编程者具有生物学的思考方法，入门难度大一点。但一旦入了门，就可得到广泛应用。由于这种方法编程时无须对角色的活动规律做详细规定，应用于复杂问题，通常会比前一种方法更省力。

AI 具有自主性、自适应性、智能交互、大数据处理能力、学习能力、实时响应、高度集成、模式识别、错误容忍性、并行处理能力等几个主要特点，伴随技术持续迭代升级，AI 系统的功能与综合能力将持续拓展、优化。

二、人工智能的术语

（一）数据

数据是人工智能的“燃料”，是实现智能决策的基础。数据的质量（如准确性、相关性、完整性、一致性）和数量（如规模、覆盖范围、多样性）直接影响人工智能系统的性能和智能化程度。例如，在食堂的智能推荐系统中，其有效运行依赖于大量高质量的数据。这些数据包括但不限于：学生的历史消费记录（如菜品、时间、频率）、明确记录或通过行为分析推断出的口味偏好（如对辣度、甜度、特定食材的偏好或规避），以及潜在的营养需求信息。这

些数据是系统进行个性化推荐、预测菜品需求，优化采购与备餐计划的基本依据。消费记录、学生口味偏好等数据是系统进行个性化推荐的重要依据。

（二）算力

算力是指计算机系统的计算能力，是人工智能的“引擎”，为人工智能算法的运行提供了必要的计算资源。强大的算力能够支持复杂的算法运行，加速模型训练和推理过程。例如，在停车场空位识别系统中，需要实时处理来自多个摄像头采集的视频或图像数据，通过图像处理和计算机视觉算法分析每个车位的占用情况，如图 1-1 所示，该场景需要充足的计算能力以保证信息更新的及时性和准确性。若缺乏足够的边缘计算单元或云端计算资源支持，系统的实时性和实用性将大打折扣。因此，充足的算力是确保此类系统高效、稳定运行的物理基础。



图 1-1　车位占用状态智能监测示意图

（三）算法

算法是人工智能的“大脑”，是智能决策的核心逻辑。算法规定机器从数据中学习、推理、决策的完整逻辑路径：机器依托训练数据提取特征、挖掘内在规律，可实现预测、分类、聚类、规划、优化等功能。以图书馆座位管理系统为例，核心为一套座位分配调度算法：实时采集座位传感数据，结合学生预约记录、历史使用习惯、区域座位热度动态调配座位资源，引导学生前往空位、处理预约需求，管控长期闲置占位，依托算法智能调度提升场馆资源利用率与师生使用体验，完成动态座位分配与优先级调度。

三、人工智能的分类

人工智能可以依据应用场景、智能层级划分不同类型，各类别便于区分 AI 落地场景与技术特征。

（一）按用途进行划分

1. 决策式人工智能

核心功能为场景分析与自主决策。它通过评估多种选项和可能的结果，帮助用户或系统选择最佳的行动方案。例如，自动驾驶场景中，车辆依靠决策式人工智能系统判定加速、减速、变道的时机。

2. 生成式人工智能

核心能力为全新内容创作，依托训练数据自动产出文本、图像、音频等素材。使用者可上传多篇论文至生成式 AI，模型可整合文稿核心观点与关键结论，辅助生成文献综述初稿。

（二）按智能水平进行划分

1. 弱人工智能（Weak AI）

也称为狭义人工智能（ANI），专注于特定任务或领域。弱人工智能系统主要执行特定类型功能，如语音识别、图像识别或自然语言处理等。它们通过大量数据和算法训练来优化特定任务，从而提高效率和准确性。尽管在特定领域表现出色，但弱人工智能缺乏对其他领域的适应能力和自我学习能力。

2. 强人工智能（Strong AI）

也称为广义人工智能（AGI），旨在模仿人类的全面智能。强人工智能系统能够像人类一样进行感知、理解、推理、决策和学习，并可以在任何领域或任务中表现出与人类相当的智能。它们具备自我意识和情感，能够在没有明确编程的情况下适应新环境和任务。强人工智能目前仍处于研究和开发阶段，其实现需要大量的计算资源和先进的技术。

3. 超人工智能（Superintelligent AI）

是指机器或系统在所有领域和任务中都具备超越人类的智能。超人工智能作为一种思想设想，可通过自我学习和进化，不断优化和扩展其智能水平。超人工智能不仅具备全面的智能，还具有高度的创造性和自主性，能够进行独立思考和创新。然而，超人工智能目前仍处于理论阶段，其实现需要克服许多技术和伦理挑战。

在实际应用中，弱人工智能是比较常见的应用形式。我们日常生活中使用的许多智能助手、语音识别软件和推荐系统等都是基于弱人工智能技术实现的。强人工智能的应用相对较少，但随着技术的不断进步，越来越多的领域开始探索强人工智能的应用潜力，如医疗、金融和交通等。超人工智能目前仍处于研究和探索阶段，但其潜在的应用前景令人充满想象。

四、人工智能的特点

人工智能（AI）核心特征包含智能化、学习能力、自主性三大维度，伴随技术持续迭代，AI 系统功能与综合能力将持续优化延伸。

（一）智能化

智能化是 AI 最核心特征。机器模拟人类思考、决策逻辑，可仿照人类思维模式分析、处

理问题。以机器学习为例，通过大数据和算法的应用，人工智能系统可以自动从数据中学习，不断提升自身的智能水平。这种智能化使得人工智能在多个领域具有广泛的应用前景。

（二）学习能力

人工智能系统的学习能力是其另一个重要特点。通过不断接收和处理新的信息，人工智能系统可以从中获取知识和经验，并运用于问题的解决中。不仅如此，人工智能系统还能够通过与环境的交互，实现对自身的学习和进化。通过反馈机制和自我调整算法，人工智能系统可以根据实际情况进行优化和改进，不断提高自身的性能和准确性。

（三）自主性

人工智能系统的自主性是指其独立进行决策和行动的能力。相对于传统的预设规则和指令，人工智能系统能够根据自身的智能和学习能力，主动进行决策和行动，而无需人为干预。

五、人工智能技术架构

人工智能技术架构通常由基础层、技术层和应用层构成，如图 1-2 所示。



图 1-2　人工智能技术架构

（一）基础层

基础层是人工智能技术的基础，负责数据的收集和处理。它包括数据、平台系统和硬件设施。数据部分涵盖图像、视频、语音、文字和结构化数据，这些是人工智能进行学习和决策的基础。平台系统包括云计算、云操作系统、大数据系统和计算资源调度，这些系统为数据的处理和分析提供了强大的支持。硬件设施如芯片、云计算服务器、传感器、GPU/CPU 和 RAM，为人工智能运行提供必要的物理基础。

（二）技术层

技术层是人工智能技术的核心，它包括通用技术、算法和底层框架。通用技术如自然语言处理、计算机视觉和智能语音，为人工智能提供了基础能力。算法部分则包含深度学习、强化学习、集成学习、回归模型、决策树模型、支持向量机和聚类分析等，这些算法是实现人工智能功能的关键。底层框架如 Tensorflow、PyTorch、Caffe 和 MXNet，为算法的实现提供了平台和工具。

（三）应用层

应用层是人工智能技术在各个行业和领域中的具体实现。它涵盖了广泛的解决方案，包括智能工业、智能交通、智慧城市、智能医疗、自动驾驶、智能农业、智能政务、智慧财务和智能零售。这些解决方案利用人工智能技术，通过感知和决策能力，提高各行业的效率和智能化程度。

这三个层次共同构成了人工智能技术架构，从底层的数据收集和处理，到中层的算法和框架，再到顶层的具体应用，形成了一个完整的生态系统。这种结构不仅确保了人工智能系统的高效运行，也为未来的技术创新和应用拓展提供了广阔的空间。

任务二　人工智能技术基础

一、机器学习

机器学习（Machine Learning，ML）是实现人工智能的核心技术之一，是让计算机系统具备“学习”能力的关键方法。简单来说，机器学习就是让机器从数据中自动“学习”规律和模式，并利用习得知识对新场景做出预测或决策，而无需人工编写面面俱到的固定程序。

（一）机器学习的基本原理

机器学习的运作方式与人类的学习过程有相似之处。人类通过观察大量实例、总结经验来提高认知和决策能力。同样，机器学习系统也需要“喂食”大量的数据，这些数据就是机器学习的“燃料”。通过特定的算法（学习规则、数学模型），机器能够从这些数据中自动发现隐藏的规律、特征和关联。

这个过程可以分为两个主要阶段。

1. 训练阶段

将有已知结果的数据输入算法中，算法通过不断调整其内部参数，来学习数据中的模式。例如，给系统提供成千上万张标注了“猫”或“狗”的图片，算法会学习区分猫和狗的特征。

2. 推理/预测阶段

当模型训练完成后，就可以用它来处理新的、未见过的数据。例如，给训练好的模型一张新的动物图片，它就能根据学到的知识判断出这是一只“猫”还是一只“狗”。

（二）机器学习的主要类型

根据学习方式的不同，机器学习主要可以分为以下几种类型。

1. 监督学习

这是最常见的学习方式。在训练数据中，每个样本都带有明确的“答案”。学习的目标是建立一个从输入到输出的映射关系。常见的应用包括：

● 分类：预测离散的类别。例如，判断一封邮件是“垃圾邮件”还是“正常邮件”，或识别图像中的物体。

● 回归：预测连续的数值。例如，根据房屋的面积、位置等信息预测房价。

2. 无监督学习

训练数据没有标签，算法需要自己发现数据的内在结构或模式。常见的应用包括：

● 聚类：将相似的数据点分组。例如，根据用户的消费行为将客户分成不同的群体，以便进行精准营销。

● 降维：简化数据，保留最重要的信息，常用于数据可视化或作为其他算法的预处理步骤。

3. 强化学习

这是一种通过“试错”和“奖励”机制来学习的方法。智能体（Agent）在与环境的交互中，根据其采取的行动所获得的奖励或惩罚来调整策略，目标是最大化长期累积奖励。训练逻辑类似动物正向激励训练，正确动作获得正向反馈，对偏差动作施加负向约束。强化学习在自动驾驶、游戏 AI（如 AlphaGo）等领域有重要应用。

（三）机器学习与人工智能的关系

在人工智能的技术架构中，机器学习算法构成了技术层的核心。它依赖于基础层提供的海量数据和强大算力，并最终支撑起应用层各类智能服务落地。可以说，没有机器学习，人工智能就难以实现从数据到智能的转化。

二、深度学习

深度学习是机器学习的一个重要分支，也是当前推动人工智能取得突破性进展的核心技术。它通过模拟人脑神经元的工作方式，构建具有多层结构的人工神经网络，能够自动从海量数据中逐层提取和学习高阶抽象特征，从而解决图像识别、语音处理等复杂问题。

（一）深度学习的基本原理

深度学习的灵感来源于生物神经网络。人脑由数百亿个神经元相互连接而成，通过电信

号传递信息并进行复杂的信息处理。深度学习中的“神经网络”是对这一结构的数学模拟。一个典型的深度神经网络由多个层次构成：

● 输入层：接收原始数据，如图像的像素值、声音的波形数据或文本的字符编码。

● 隐藏层：这是深度学习的“核心”。一个网络通常包含多个隐藏层（“深度”即指隐藏层的数量）。每一层都会对前一层的输出进行非线性变换，提取出更高级的特征。例如，在图像识别中，第一层可能学习到边缘和线条，第二层学习到角点和简单形状，更深层则能组合出眼睛、鼻子等复杂部件，最终识别出整张人脸。

● 输出层：产生最终的预测或分类结果，如判断图片中的物体是“猫”还是“狗”。

（二）深度学习与机器学习的关系

深度学习是机器学习的一种特定方法。可以将机器学习看作一个“大伞”，而深度学习是其中一种非常强大且高效的“工具”。

1. 传统机器学习

通常需要人工进行特征工程。这意味着专家需要根据问题领域，手动设计和提取数据的关键特征。例如，要识别手写数字，专家可能需要手动编写程序来提取数字的笔画、闭合区域等特征，然后将这些特征输入 SVM（支持向量机）或决策树等算法进行分类。这种方法依赖于专家的知识，且在处理高维、非结构化数据（如图像、音频）时效果有限。

2. 深度学习

其最大的优势在于能够自动进行特征学习。它直接将原始数据（如整个图像）输入网络，由网络的多层结构自动发现和学习最优的特征表示，大幅降低人工特征工程依赖，可直接处理原生、高复杂度原始素材。

深度学习和机器学习、人工智能的关系如图 1-3 所示。深度学习的核心是神经网络模型，使用具有多层非线性处理单元的神经网络来对大量数据进行建模和学习。与传统机器学习算法相比，深度学习具有更强的表达能力和学习能力，可以更好地处理大规模和高维度数据，因此在计算机视觉、自然语言处理和语音识别等领域应用广泛。

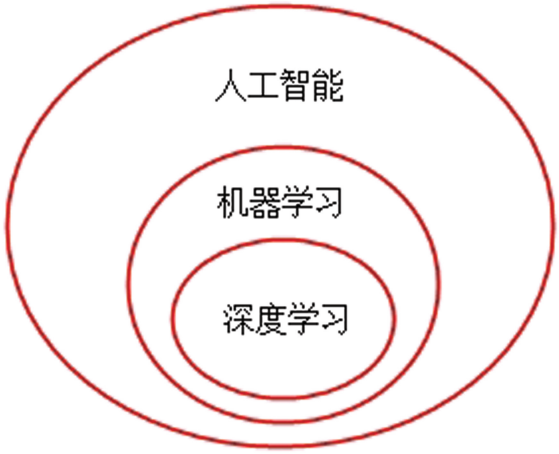


图 1-3　深度学习、机器学习与人工智能的包含关系

三、计算机视觉

计算机视觉是人工智能领域中一个重要的分支，它致力于让计算机具备“看”和“理解”图像与视频的能力，模拟人类的视觉系统，使机器能够从视觉信息中提取意义并做出相应决策。简单来说，计算机视觉就是让机器拥有“眼睛”和“大脑”，能够识别、分析和理解视觉世界。

（一）计算机视觉的基本原理

计算机视觉的核心任务是将二维图像或视频数据转化为有意义的三维世界信息。这一过程涉及多个步骤。

1. 图像获取

通过摄像头、扫描仪等设备捕获原始图像数据。这些数据以像素矩阵的形式存储，每个像素包含颜色和亮度信息。

2. 预处理

对原始图像进行去噪、增强、归一化等操作，提高图像质量，为后续分析做准备。

3. 特征提取

识别图像中的关键特征，如边缘、角点、纹理等。传统方法依赖于人工设计的特征提取器（如 SIFT、HOG），而现代方法则多采用深度学习自动提取特征。

4. 高级理解

基于提取的特征进行更高层次的分析，如目标识别、场景理解、行为分析等。

计算机视觉系统通过这些步骤，能够实现从简单到复杂的视觉任务，从识别图像中的物体到理解整个场景的语义。

（二）计算机视觉的主要任务

计算机视觉涵盖多种任务类型，每种任务解决不同的视觉理解问题：

1. 图像分类

确定整张图像所属的类别。例如，判断一张图片中是“猫”还是“狗”。这是计算机视觉中最基础的任务。

2. 目标检测

不仅识别图像中包含的物体类别，还要确定每个物体在图像中的具体位置（通常用边界框表示）。例如，在自动驾驶系统中识别道路上的行人、车辆和交通标志。

3. 语义分割

将图像中的每个像素分类到特定的语义类别中，实现像素级别的理解。例如，将医学影像中的不同组织区域精确区分。

4. 实例分割

在语义分割的基础上，进一步区分同一类别的不同个体。例如，区分图像中多个人物、多个同类物体的具体轮廓。

5. 姿态估计

检测人体或物体的关键点位置，用于分析姿势和动作。这在运动分析、虚拟现实等领域有广泛应用。

6. 图像生成与编辑

创建全新图像或修改已有图像，生成对抗网络（GAN）在该领域成果突出，广泛应用于艺术创作、图像修复等场景。

（三）计算机视觉与深度学习的关系

深度学习，特别是卷积神经网络（CNN），已成为计算机视觉领域的核心技术。在“人工智能的技术架构”中，CNN 属于技术层的关键算法。

1. CNN 的工作原理

CNN 通过模拟生物视觉皮层的工作方式，使用多层卷积核自动提取图像的层次化特征。浅层网络学习边缘和纹理等底层特征，深层网络则组合这些特征形成更高级的语义表示。

2. 端到端学习

与传统计算机视觉方法需要人工设计特征不同，深度学习方法可以直接从原始像素学习到最终任务，实现端到端的训练，大大提高了性能和效率。

3. 迁移学习

计算机视觉领域广泛使用迁移学习技术，将在一个大型数据集（如 ImageNet）上预训练的模型应用到特定任务中，即使目标领域的数据量有限也能取得良好效果。

计算机视觉作为人工智能感知世界的重要媒介，其技术进步将持续推动智能系统在各个领域的应用落地，为人类社会带来更智能、更便捷的生活体验。

四、自然语言处理

（一）自然语言处理概述

自然语言处理（Natural Language Processing，以下简称 NLP）是人工智能重要分支，研究依托计算机建模解析人类自然语言的学科，实现对文字和语音在形、音、义全维度的处理与识别，典型应用覆盖机器翻译、文本摘要自动生成、文本分类、语音合成、情感分析等。

自然语言处理的发展历程如下：

1. 早期探索阶段（20 世纪 50 年代至 70 年代）

自然语言处理的起源可以追溯到 20 世纪 50 年代，当时计算机科学家开始尝试让计算机处理和翻译自然语言。1954 年，IBM 与乔治城大学（Georgetown University）合作开展早期

机器翻译实验，尝试将俄语句子译为英语，本次实验效果有限，但拉开了自然语言处理研究的序幕。

2. 统计方法兴起阶段（20 世纪 80 年代至 90 年代）

20 世纪 80 年代，随着统计学和机器学习理论的发展，统计方法开始被引入自然语言处理。例如，隐马尔可夫模型被用于语音识别和词性标注。统计方法通过分析大量标注数据来学习语言的模式和规律，相比基于规则的方法，其灵活性和适应性更强。在这一时期，自然语言处理在一些经典任务上取得了显著进展，如机器翻译、词性标注、句法分析等。例如，基于统计的机器翻译系统通过分析双语语料库来学习翻译规则。

3. 深度学习推动阶段（21 世纪初至今）

2010 年前后，深度学习在 NLP 领域快速发展。2013 年 Word2Vec 问世，革新词嵌入技术：依托无监督学习把词汇映射至高维向量空间，让语义相近词语的向量距离更近。此后循环神经网络、LSTM、GRU 被大量用于语言建模、机器翻译、文本生成等序列任务。2017 年 Transformer 架构问世，开启 NLP 全新发展阶段。现阶段预训练模型成为行业主流，BERT、GPT、RoBERTa、T5 等模型在文本分类、实体识别、问答、机器翻译等全类 NLP 任务中性能突出。

（二）自然语言处理技术

1. 数据预处理

数据预处理是自然语言处理的第一步，其目的是将原始文本数据转换为适合模型处理的格式。预处理的质量直接影响后续模型的性能。

（1）文本清洗

①去除噪声。删除文本中的无关字符，如 HTML 标签、特殊符号、冗余空格等。例如，对于从网页爬取的文本，需要去除<html><body>等标签，只保留文本内容。

②统一格式。将文本中的字符统一为小写或大写，以减少词汇的多样性。例如，将“Apple”和“apple”统一为“apple”，这样可以避免模型将它们视为两个不同的单词。

（2）分词

将文本分割成单词、词组或句子等单元。分词是自然语言处理的基础，因为大多数语言模型都是基于单词或词组进行处理的。

①中文分词：中文是以字为基本单位的，没有明显的单词分隔符。因此，中文分词是一个复杂的过程。例如，“我爱自然语言处理”可以分割为“我/爱/自然语言处理”也可以分割为“我/爱/自然/语言/处理”。常用的中文分词算法包括基于规则的分词、基于统计的分词和基于深度学习的分词。基于规则的分词是通过人工编写的词典和规则来分割文本，例如，将“自然语言处理”作为一个整体词组。基于统计的分词是通过分析大量文本数据，计算词的出现频率和上下文信息来确定分词结果。基于深度学习的分词是利用神经网络模型自动学习分词规则，能够更好地处理歧义和新词。

②英文分词：英文单词之间有明显的空格分隔，分词相对简单。但是，英文中也存在一些特殊情况，如缩写词（“can't”）、连字符词（“well-known”）等，需要进行特殊处理。

（3）停用词删除

停用词是指在文本中频繁出现但对语义贡献不大的词汇，如“的”“和”“是”“在”等。这些词在文本中出现的频率很高，但在大多数自然语言处理任务中，它们的语义信息较少，因此可以删除以减少数据量和提高模型性能。去除停用词可以降低文本的维度，减少模型的计算量，同时提高模型对关键信息的关注度。例如，在文本分类任务中，去掉停用词后，模型可以更专注于与类别相关的词汇。

（4）词干提取和词形还原

①词干提取：将单词还原到其基本形式，通常是通过简单的规则来去掉单词的后缀。例如，将“running”还原为词干“run”，将“happiness”提取词干为“happi”。词干提取的算法有 Porter Stemmer、Lancaster Stemmer 等。这些算法基于一系列的规则来逐步去掉单词的后缀，但可能会产生一些不准确的结果，因为它们没有考虑单词的语义和上下文。

②词形还原：与词干提取类似，但词形还原会考虑单词的词性，将单词还原到其基本词形。例如，“running”还原为“run”，“was”还原为“be”。词形还原通常使用词典和词法规则来实现，比词干提取更准确，但计算量也更大。在一些对语义要求较高的任务中，如文本相似度计算，词形还原比词干提取更有优势。

2. 语言模型

语言模型是自然语言处理的核心，用于计算句子在语言中出现的概率。它可以帮助计算机判断一个句子是否符合语言的语法和语义规则，从而实现文本生成、机器翻译等功能。

（1）基于规则的语言模型

通过人工定义的规则来分析语言结构。这些规则通常是基于语言学家对语言的语法和语义的研究成果。例如，定义主语＋谓语＋宾语的句子结构规则，或者定义名词短语、动词短语的构成规则。基于规则的语言模型依赖于人工编写的规则，这些规则很难覆盖语言的所有情况，尤其是对于复杂的语言现象和新出现的语言表达。此外，语言规则的编写和维护成本较高，且难以适应不同语言和方言的差异。

（2）基于统计的语言模型

通过分析大量文本数据来统计词汇之间的关系。最经典的统计语言模型是 N 元语法模型。以二元语法模型为例，它计算一个词出现的概率是基于它前面一个词的概率。例如，在句子“我爱自然语言处理”中，计算“爱”出现的概率是基于它前面的词“我”。“自然语言处理”出现的概率是基于它前面的词“爱”。N 元语法模型的阶数越高，能够捕捉的语言结构越复杂，但计算量也越大，且数据稀疏问题越严重。

（3）Transformer 架构与预训练语言模型

Transformer 是一种基于自注意力机制的神经网络架构，它能够捕获输入与输出之间的全局依赖关系。Transformer 架构的核心是自注意力机制，它能够计算输入序列中每个词与其他词之间的关系权重，从而实现对全局信息的建模。例如，在句子“我爱自然语言处理”中，自注意力机制可以计算“我”与“爱”“自然语言处理”之间的关系权重，以及“爱”与“我”“自然语言处理”之间的关系权重等。通过这种方式，Transformer 能够更好地理解句子的语义结构。

预训练语言模型是基于 Transformer 架构的一种创新应用，它通过在大规模无监督文本数据上进行预训练，学习语言的通用特征，然后在特定任务上进行微调。预训练语言模型的出现极大地推动了自然语言处理的发展。

3. 词嵌入

词嵌入是自然语言处理中的一个重要技术，它将词语映射到高维向量空间中，使得具有相似语义的词在向量空间中距离较近。词嵌入技术为自然语言处理提供了一种有效的表示方法，使得计算机能够更好地理解和处理语言。

（1）传统词嵌入方法

One-Hot 编码是一种简单的词嵌入方法，它将每个词表示为一个独热向量。假设词汇表大小为 N，那么每个词对应的独热向量是一个长度为 N 的向量，其中只有一个元素为 1，其余元素为 0。例如，词汇表为｛“我”，“爱”，“自然”，“语言”，“处理”｝，那么“我”的独热向量为［1，0，0，0，0］，“爱”的独热向量为［0，1，0，0，0］。One-Hot 编码的优点是简单直观，但它存在很大的缺点。首先，它没有考虑词与词之间的语义关系，所有词的向量都是正交的，无法体现词的相似性。其次，随着词汇表的增大，One-Hot 编码的维度会急剧增加，导致计算量和存储量都非常大。

（2）基于神经网络的词嵌入方法

Word2Vec 是一种基于神经网络的词嵌入模型，它通过上下文信息来学习词向量。Word2Vec 有两种训练方式：CBOW（Continuous Bag-of-Words）和 Skip-Gram。CBOW 是根据上下文词来预测目标词，例如，给定上下文词“我”和“自然语言”，预测目标词“爱”。Skip-Gram 是根据目标词来预测上下文词，例如，给定目标词“爱”，预测上下文词“我”和“自然语言”。Word2Vec 通过训练得到的词向量能够很好地捕捉词与词之间的语义关系。例如，在向量空间中，“国王”和“王后”的向量距离较近，“男人”和“女人”的向量距离较近。Word2Vec 的一个重要特点是它能够进行类比推理，例如，“国王”－“男人”＋“女人”≈“王后”。这表明 Word2Vec 学习到了词与词之间的语义关系和类比关系。

（3）上下文化词嵌入

上下文化词嵌入是指词的表示不仅取决于词本身，还取决于它所在的上下文。例如，“苹果”在“苹果是一种水果”中表示一种水果，在“苹果公司是一家科技公司”中表示一家公司。上下文化词嵌入能够更好地捕捉词的语义多样性。

4. 语法和语义分析

语法和语义分析是自然语言处理中的两个重要环节，它们分别从语法结构和语义内容两个方面对文本进行分析和理解。

（1）语法分析

①短语结构分析：短语结构分析是将句子分解为短语结构树的过程。短语结构树是一种层次化的结构，它表示句子的语法成分和它们之间的关系。例如，句子“我爱自然语言处理”可以分解为短语结构树：｛S（NP 我）［VP 爱（NP 自然语言处理）］｝。在这个短语结构树中，“S”表示句子，“NP”表示名词短语，“VP”表示动词短语。短语结构分析可以帮助计算机理解句子的语法结构，例如，确定句子的主语、谓语和宾语。

②依存句法分析：依存句法分析是分析句子中词与词之间的依存关系的过程。依存关系是指一个词依赖于另一个词，这种关系反映了句子的语法结构和语义结构。例如，在句子“我爱自然语言处理”中，“爱”是中心词，“我”和“自然语言处理”都依赖于“爱”。依存句法分析可以帮助计算机理解句子的语义结构，例如，确定句子中的主谓关系、动宾关系等。依存句法分析通常使用基于图的方法或基于转移的方法。基于图的方法是将句子表示为一个图，图中的节点表示词，边表示依存关系。基于转移的方法通过一系列转移操作构建依存关系树，该方法计算效率较高，模型效果随训练方案、数据集不同存在差异。

（2）语义分析

①词义消歧：是指确定一个多义词在具体语境中的具体含义。例如，“苹果”可以表示一种水果，也可以表示一家科技公司。在句子“苹果是一种水果”中，“苹果”的含义是水果，在句子“苹果公司是一家科技公司”中，“苹果”的含义是科技公司。词义消歧是自然语言处理中的一个难题，因为它需要计算机理解上下文信息。

②指代消解：是指确定文本中的代词或名词短语所指代的具体对象。例如，在句子“他喜欢自然语言处理”中，“他”指代的是谁？指代消解可以帮助计算机更好地理解文本的语义结构，例如，确定文本中的主语和宾语。

5. 自然语言理解和自然语言生成

自然语言理解和自然语言生成是自然语言处理的两个重要分支，它们分别负责让计算机理解人类语言和生成人类语言。

（1）自然语言理解

自然语言理解是让计算机理解人类语言的细微之处，包括上下文、意图和情感。它可以帮助计算机正确解析用户输入的文本，从而实现智能交互。

自然语言理解的主要任务包括意图识别、槽位填充和语义解析。

①意图识别是指确定用户输入的文本所表达的意图，例如，在用户输入“我想订一张明天去北京的机票”时，意图识别系统需要识别出用户的意图是订机票。

②槽位填充是指提取用户输入文本中的关键信息，例如，在订机票的例子中，槽位填充系统需要提取出发日期（明天）、目的地（北京）等关键信息。

③语义解析是指将用户输入的文本转换为计算机能够理解的语义表示，例如，将“我想订一张明天去北京的机票”转换为一个语义结构，其中包含用户的意图、出发日期、目的地等信息。

（2）自然语言生成

自然语言生成是基于规则集或数据库生成可被人类理解的文本。它可以帮助计算机将数据或信息转换为自然语言文本，从而实现智能交互。

自然语言生成的主要任务包括文本摘要、机器翻译、问答系统和文本生成。

①文本摘要是指从长文本中提取关键信息，生成简短的摘要文本。

②机器翻译是指将文本从一种语言翻译成另一种语言。

③问答系统是指根据用户的问题生成相应的答案。

④文本生成是指根据给定的提示或主题生成自然语言文本。

五、知识图谱

知识图谱本质上是结构化的语义知识库，是一种由节点和边组成的图数据结构，以符号形式描述物理世界中的概念及其相互关系，其基本组成单位是“实体—关系—实体”三元组，以及实体及其相关“属性—值”。不同实体之间通过关系相互联结，构成网状的知识结构。

知识图谱可用于反欺诈、不一致性验证、组团欺诈等公共安全保障领域，需要用到异常分析、静态分析、动态分析等数据挖掘方法。特别值得一提的是，知识图谱在搜索引擎、可视化展示和精准营销方面有很大的优势，已成为业界的热门工具。但是，知识图谱发展仍面临严峻挑战，例如数据噪声，即原始数据存在错误、冗余等问题。随着知识图谱应用持续落地，仍有大量关键技术亟待突破。

任务三 人工智能大模型

一、盘古大模型

盘古大模型是华为推出的盘古系列 AI 大模型，包括盘古 NLP 大模型、盘古 CV 大模型、盘古气象大模型。

（一）盘古 NLP 大模型

盘古 NLP 大模型可用于内容生成、内容理解等方面，并首次使用 Encoder-Decoder 架构，兼顾 NLP 大模型的理解能力和生成能力，保证了模型在不同系统中的嵌入灵活性。在下游应用中，仅需少量样本和可学习参数即可完成千亿规模大模型的快速微调和下游适配。在中文语言理解评测基准 CLUE 榜单中，盘古 NLP 大模型在总排行榜及分类、阅读理解单项均表现优异，刷新多项榜单纪录；总排行榜得分 83.046，多项子任务得分业界领先，是同期测评中距离人类理解水平（85.61）较近的预训练模型之一。

（二）盘古 CV 大模型

盘古 CV 大模型支持分类、分割、检测任务，也是首次实现模型按需抽取的业界最大 CV 大模型，同时兼顾判别与生成能力。基于模型大小和运行速度需求，自适应抽取不同规模模型，AI 应用开发快速落地。使用层次化语义对齐和语义调整算法，依托层级化语义对齐与调优算法提升浅层特征区分度，小样本学习能力实现行业顶尖水平。

（三）盘古气象大模型

盘古气象大模型实现气象预报精度首次超过传统数值方法，推演速度提升万倍以上，提供秒级天气预报，可对位势高度、湿度、风速、温度、气压等变量进行 1 小时到 7 天的精准

预测。借助创新的 3DEST 网络结构以及分层时间聚合算法，该模型在气象预报的关键要素和常用时间范围上精度均超过传统预报方法。

二、豆包大模型

豆包大模型是字节跳动推出的豆包人工智能助手所采用的大模型技术。字节跳动的豆包大模型家族包含以下几种模型。

（一）豆包通用模型

豆包通用模型是字节跳动自研大语言模型（Large Language Model，LLM）专业版，支持 128K 长文本，全系列可精调，具备更强的理解、生成、逻辑等综合能力，适配问答、总结、创作、分类等多元场景。

（二）豆包通用模型 lite

字节跳动自研 LLM 模型轻量版，相较专业版，Token（令牌，是大模型处理文字时的基础单元）具备更低成本、更低延迟，为企业提供灵活经济的模型选择。

（三）豆包 · 角色扮演模型

个性化的角色创作能力，更强的上下文感知和剧情推动能力，满足灵活的角色扮演需求。

（四）豆包 · 语音合成模型

提供自然生动的语音合成能力，善于表达多种情绪，演绎多种场景。

（五）豆包 · 声音复刻模型

5 秒即可实现声音 1:1 克隆，对音色相似度和自然度进行高度还原，支持声音的跨语种迁移。

（六）豆包 · 语音识别模型

更高的准确率及灵敏度，更低的语音识别延迟，支持多语种的正确识别。

（七）豆包 · 文生图模型

更精准的文字理解能力，图文匹配更准确，画面效果更优美，擅长对中国文化元素的创作。

（八）豆包 · Function call 模型

提供更加准确的功能识别和参数抽取能力，适合复杂工具调用的场景。

（九）豆包 · 向量化模型

聚焦向量检索的使用场景，为 LLM 知识库提供核心理解能力，支持多语言。

三、讯飞星火大模型

讯飞星火大模型是科大讯飞推出的新一代认知智能大语言模型，其核心能力涵盖了文本生成、语言理解、知识问答、逻辑推理、数学能力、代码能力和多模态能力七大方向，具体介绍如下。

（一）文本生成

文本生成是讯飞星火大模型的一项核心能力，它能够通过自然语言处理技术，自动产生高质量的文本内容。这一功能在多个应用场景中展现出了巨大的价值，包括文档创作、多语种支持、图文匹配以及演讲准备等。

1. 文档一键生成

用户只需提供主题或关键词，系统便能迅速产出一篇结构完整、内容丰富的文章草稿。这项功能极大地简化了写作流程，尤其适合于需要快速撰写报告、论文或是博客文章的用户。无论是学术研究还是商业写作，都能从中获得显著的效率提升。

2. AI 撰写助手

作为写作过程中的得力伙伴，AI 撰写助手不仅可以帮助作者寻找灵感，还能对已有的文字进行优化润色。例如，当遇到难以下笔的情况时，它可以提供段落构思建议；而对于已经完成的部分，则可以通过智能算法分析并提出改进意见，使最终成品更加流畅易读。

3. 多语种文档生成

为了满足全球化交流的需求，该平台特别强化了其跨语言服务能力。除了能够在不同语言间实现精准翻译之外，还可以直接用目标语言创建新的内容。这意味着无论您的目标受众位于何处，都能够轻松制作出符合当地文化习惯的信息材料。

4. 自动配图

良好的视觉元素往往能让文本更具吸引力。基于该理念，讯飞星火内置图像搜索引擎，可依据文章内容自动匹配配图。使用该功能时，需要注意图片版权问题，商用场景需要取得图片授权。

5. 模板选择

为了让每个人都能轻松上手使用这款工具，我们提供了多种类型的文档模板供选择。无论是正式的商业计划书还是个人日记记录，都可以找到相应的样式来快速开始工作。这些预设格式经过精心设计，旨在帮助使用者节省时间的同时保持专业性。

6. 演讲备注

对于经常需要在公众面前发言的人来说，事先准备好详尽的讲稿是非常重要的。为此，讯飞星火允许用户在其提供的文本基础上添加个性化注释和提示，以便于记忆关键点并在适当时候引用。此外，还可以根据具体场景调整语言风格，确保每次演讲都能达到最佳效果。

（二）语言理解与知识问答

依托 NLP 技术深度挖掘用户提问意图，针对生活常识、医学科普、政策规范类问题给出严谨答复，可结合上下文优化答案精度并完整推导依据逻辑。

（三）逻辑推理与科研辅助

可拆解问题背景与前置假设，依托逻辑规则推导可靠结论、输出创新观点；面向科研场景，能依托已有数据完成趋势预判、假设验证，加快科研推进节奏。

（四）数学问题解决

擅长处理各类复杂的数学题目，如方程求解、立体几何图形分析、微积分运算以及概率统计等，还可为求解的题目输出完整解题步骤，帮助学生理解相关知识点。

（五）编程支持

不但可以根据函数名或注释智能生成代码片段，减轻开发者负担，还可以对现有代码进行检查，指出可能存在的语法错误或逻辑漏洞，并提出改进建议。

（六）多模态交互体验

多模态交互是指通过结合多种感官模式（如视觉、听觉等）来实现人与机器之间的自然沟通。在人工智能领域，这种技术尤为重要，因为它使得用户能够以更直观、更人性化的方式与智能系统进行互动。讯飞星火大模型在这方面表现出色，不仅支持文本和语音输入输出，还能处理图像、视频等多种类型的数据，从而提供更加丰富多样的用户体验。

1. 图像识别与描述

用户可以上传图片，让 AI 分析并生成详细的文字说明。这对于教育、艺术鉴赏或是简单的信息检索都非常有用。例如，学生可以通过拍摄书本上的概念图来获取相关知识点解释；艺术家则可以利用这项功能为自己的作品添加描述性标签。

2. 基于图片的内容问答

除了基本的图片理解外，该系统还能够根据图片内容回答具体问题。比如，如果一张图片展示了某种植物，你可以询问它的名字或者生长习性等信息。这极大地扩展了传统搜索引擎的功能范围。

3. 音频/视频生成

依据用户提供的文字脚本或指令，自动合成相应的声音效果或创建简短的视频片段。这对于内容创作者来说，这是一项强大的工具，可以帮助他们快速制作出具有专业水准的作品，无论是用于社交媒体分享还是正式场合展示都非常合适。

四、文心大模型

文心大模型是百度推出的自然语言处理预训练模型系列，以“文心一言”为代表，它有

多个版本，每个版本都在前一个版本的基础上进行了改进和优化。

（一）文心一言 3.5

2023 年 8 月 31 日，文心一言 3.5 面向全社会开放，相比 3.0 版本在推理能力、代码写作能力等方面有显著提升。该版本在自然语言处理领域表现出了强大的性能，包含了基础的文本理解、生成、分类等任务的能力。

（二）文心一言 4.0

2023 年 10 月 17 日，百度正式发布文心一言 4.0，多项综合能力对标 GPT-4。与 3.5 版本相比，4.0 版本在模型结构、训练数据、算法优化等方面都有所提升，在复杂指令理解、专业领域问答和内容创作等方面表现更为出色。

（三）文心一言 4.5

2025 年 3 月 16 日，百度推出文心一言 4.5 版本，在生成速度和效果上都有了进一步提升。这一版本在保持高质量输出的同时，显著提高了响应速度。此外，百度还基于文心大模型推出了多个行业应用，如面向农业领域的智能解决方案，为农业生产提供精准指导和服务，助力农业智能化发展。

任务四 人工智能现状

一、人工智能的发展历程

自诞生以来，人工智能（AI）便在充满未知的道路上探索，曲折起伏，可将这段发展历程大致划分为 5 个阶段：

第一阶段：起步发展期（1956 年—20 世纪 60 年代）

1956 年夏季，美国达特茅斯学院举办学术研讨会，约翰·麦卡锡、马文·明斯基、克劳德·香农等科学家首次提出“人工智能（Artificial Intelligence）”这一术语，正式确立了人工智能作为独立学科的地位，此次会议也被公认为人工智能诞生的标志。人工智能概念提出后，发展出了符号主义、联结主义（神经网络），相继取得了一批令人瞩目的研究成果，如机器定理证明、跳棋程序、人机对话等，掀起了人工智能发展的第一个高潮。

第二阶段：反思发展期（20 世纪 70 年代）

人工智能发展初期的突破性进展大幅提升了人们对人工智能的期望，人们开始尝试更具挑战性的任务，但受限于算力不足、理论尚未成熟，早期设定的很多目标无法实现，人工智能发展进入低谷。

第三阶段：应用发展期（20 世纪 80 年代）

专家系统模拟人类专家的知识和经验解决特定领域的问题，实现了人工智能从理论研究走向实际应用、从一般推理策略探讨转向运用专门知识的重大突破。而机器学习（特别是神经网络）探索不同的学习策略和各种学习方法，在大量的实际应用中也开始慢慢复苏。

第四阶段：平稳发展期（20 世纪 90 年代—2010 年）

由于互联网技术的迅速发展，人工智能的创新研究得以加速，促使人工智能技术进一步走向实用化，人工智能相关的各个领域都取得了进步。在 2000 年初，由于专家系统的项目都需要编码太多的显式规则，这降低了效率并增加了成本，人工智能研究的重心从基于知识系统转向了机器学习方向。

第五阶段：蓬勃发展期（2011 年至今）

这一阶段以大数据、云计算与专用计算芯片（GPU/TPU）的成熟为支撑，深度学习算法的突破推动人工智能进入爆发式增长阶段。2012 年卷积神经网络（CNN）革新图像识别；AlphaGo 等案例验证了 AI 在复杂决策场景的潜力；2017 年 Transformer 架构重塑自然语言处理，催生大语言模型，多模态技术进一步拓宽应用边界。

随着大数据、云计算、互联网、物联网等信息技术的发展，在感知数据和图形处理器等计算平台推动下，以深度神经网络为代表的人工智能技术取得了飞速发展，大幅跨越了科学与应用之间的技术鸿沟，诸如图像分类、语音识别、知识问答、人机对弈、无人驾驶等人工智能技术实现了重大的技术突破，迎来爆发式增长的新高潮。

二、我国人工智能发展现状

伴随国内社会经济转型，人口红利消退、老龄化加深、高端人才缺口扩大，核心技术自主攻关成为亟待破解的课题。在此背景下，发展 AI 与智能装备可通过“机器换人”拉动产业升级，为经济、社会难题提供解决方案，“人工智能 ＋ 全行业”将成为大众创新创业的主流方向。

中国对人工智能的探索始于 20 世纪 50 年代末，当时中国科学院计算技术研究所（简称“计算所”）成立。作为国内最早的计算机研究机构之一，计算所在推动中国早期计算机科技发展的同时，也为中国在人工智能领域的研究奠定了重要基础。自那时起，随着计算机硬件和软件技术的快速发展，中国逐步积累了强大的技术实力，为后续在人工智能领域的突破和发展提供了坚实的支撑。

自 20 世纪 70 年代末起，中国学者开始接触人工智能，进入 80 年代后，国家层面正式启动人工智能相关研究计划，逐步开展系统性探索，并在语音识别、机器翻译等领域取得了初步进展。在此期间，中国计算机科学家和工程师开始积极探索人工智能的理论和技术，开展了一系列创新性的研究工作，例如，提出了基于模型的机器翻译方法、神经网络等。中华人民共和国国防科学技术工业委员会于 1984 年召开了全国智能计算机及其系统学术讨论会，1985 年又召开了全国首届第五代计算机学术研讨会。1986 年起，将智能计算机系统、智能机器人和智能信息处理等重大项目列入国家高技术研究发展计划（863 计划）。1987 年，《人工智能及其应用》一书在清华大学出版社公开出版，成为国内首部具有自主知识产权的人工智能专著。1993 年起，我国把智能控制和智能自动化等项目列入国家科技攀登计划。

国内学者在人工智能的诸多领域，如问题求解、不确定推理、泛逻辑理论、拓扑学、模式识别、图像处理、机器学习、专家系统、智能计算和智能控制等领域的基础研究也多有建树，取得了一批具有国际先进水平的创造性成果。例如，在模式识别方面，对文字识别、语音识别、指纹识别、人脸识别、虹膜识别和步态识别等进行深入研究，涉及生物医学、卫星遥感、机器人视觉、货物检测、目标跟踪、自主导航、保安、银行、交通、军事、电子商务和多媒体网络通信等应用领域。

2017 年，国务院发布了《新一代人工智能发展规划》，其中提出了大数据智能、群体智能、跨媒体智能、混合增强智能以及自主智能系统等五大智能形态，指出人工智能呈现深度学习、跨界融合、人机协同、群智开放、自主操控等新特征，标志着中国人工智能发展进入新阶段。

要实现人工智能的广泛应用和发展，关键在于技术创新和人才队伍建设。一方面，需要加大对基础科学研究的支持力度，鼓励企业和科研机构开展前沿技术探索；另一方面，还需加强高端人才的培养与引进，建立完善的教育体系以适应未来科技发展的需求。此外，政府应出台相关政策支持创新创业活动，营造良好的创新生态环境。

三、人工智能发展的启示

人工智能的目标是模拟、延伸和扩展人类智能，探寻智能本质，发展类人智能机器。其探索之路充满未知且曲折起伏。通过总结人工智能发展历程中的经验和教训，可以得到以下启示：

（一）尊重发展规律是推动学科健康发展的前提

科学技术的发展有其自身的规律，人工智能学科的发展需要基础理论、数据资源、计算平台和应用场景的协同驱动。当条件不具备时，很难实现重大突破。

（二）基础研究是学科可持续发展的基石

多伦多大学杰弗里·辛顿教授坚持研究深度神经网络数十年，奠定了人工智能蓬勃发展的重要理论基础。谷歌 DeepMind 团队长期深入研究神经科学启发的人工智能等基础问题，取得了 AlphaGo 等一系列重大成果。

（三）应用需求是科技创新的不竭之源

引领学科发展的动力主要来自科学和需求的双轮驱动。人工智能发展的驱动力除了知识与技术体系内在矛盾外，贴近应用、解决用户需求，也是创新的最大源泉。例如，人工智能专家系统实现了从理论研究走向实际应用的突破，安防监控、身份识别、无人驾驶、互联网和物联网、大数据分析等应用需求带动了人工智能的技术突破。

（四）学科交叉是创新突破的“捷径”

人工智能研究涉及信息科学、脑科学、心理科学等多个领域。20 世纪 50 年代人工智能的

出现本身就是学科交叉的结果。特别是脑认知科学与人工智能的成功结合，带来了人工智能神经网络几十年的持久发展。智能本源、意识本质等一些基本科学问题正在孕育重大突破，对人工智能学科发展具有重要促进作用。

（五）宽容失败是支持创新的题中应有之义

任何学科的发展都不可能一帆风顺，任何创新目标的实现都不会一蹴而就。人工智能七十余载的发展生动地诠释了一门学科创新发展起伏曲折的历程。可以说，没有过去发展历程中的“寒冬”，就没有今天人工智能发展新的春天。

（六）实事求是设定发展目标是制定学科发展规划的基本原则

实现全方位类人水平的机器智能是人工智能学科宏伟的终极目标，但是需要结合科技和经济社会发展水平来设定合理的阶段性研究目标，否则容易产生预期落差，进而阻碍学科发展。人工智能发展过程中的几次低谷皆因不切实际的发展目标所致。

四、学生应具备的人工智能素养

在人工智能技术深刻影响社会生产与生活的时代背景下，培养具备人工智能素养的下一代已成为教育的重要使命。人工智能素养并非要求所有学生都成为 AI 工程师，而是指学生应具备理解、应用、评估并负责任地与人工智能技术互动的综合能力。

（一）理解与使用人工智能理论和基础知识

学生应能理解人工智能的基本概念、发展历程和核心思想。这包括：

● 知晓 AI 的定义与范畴：了解人工智能是让机器模拟、延伸和扩展人类智能的技术科学，认识其在机器人、语言识别、图像识别等领域的应用。

● 掌握核心三要素：深刻理解“数据、算法、算力”作为人工智能发展的三大支柱，并能举例说明它们在实际应用（如智能推荐、自动驾驶）中的作用。

● 区分 AI 类型：能够辨析弱人工智能（如语音助手）与强人工智能（理论构想）的区别，理解当前技术主要处于弱人工智能阶段。

（二）理解与使用人工智能技术及算法思维

学生应具备运用技术思维分析和解决问题的能力，并能与 AI 技术进行基础交互。

● 发展算法与编程思维：学习基本的算法逻辑和编程思想，理解计算机如何通过指令序列解决问题，为理解 AI 的“大脑”（算法）打下基础。

● 理解关键技术原理：对机器学习、深度学习、计算机视觉、自然语言处理等核心技术有初步认知，了解它们如何让机器“学习”和“看懂/听懂”世界。

● 实践情景式解决问题：能够运用 AI 工具或思维，针对特定场景（如优化校园垃圾分类、设计智能灌溉方案）提出创新性的解决方案。

（三）发展与应用人工智能技术的能力

学生应能主动探索和应用 AI 技术，激发创新潜能。

（四）养成批判性思维与伦理道德素养

这是人工智能素养中至关重要的一环，确保技术被负责任地使用。

● 认识 AI 的社会影响：理解人工智能在带来便利的同时，也可能引发就业结构变化、数字鸿沟、信息茧房等社会问题。

● 具备批判性评估能力：不盲目相信 AI 的输出，能够质疑 AI 生成内容的真实性、客观性和潜在偏见，培养独立思考和判断能力。

五、人工智能应用

AI 与各行业深度融合，正在重构传统产业业态，目前已在金融、家居、交通、医疗、教育等场景规模化落地，深度影响社会生产与日常生活。

（一）金融

人工智能在金融领域有着广泛的应用，包括风险控制、交易分析和客户服务等方面。首先，人工智能可以通过分析大量的金融数据，帮助金融机构识别和评估风险。其次，人工智能能在股票交易（如图 1-4 所示）和外汇交易等方面提供参考性分析与趋势预判，辅助投资者理性制定决策。此外，人工智能还可以应用于金融客户服务领域，通过自然语言处理和智能机器人等技术，实现智能客服和自助银行等服务。



图 1-4　人工智能股票交易

（二）家居

智能家居是在互联网影响之下物联化的体现。智能家居通过物联网技术（手机＋APP，如图 1-5 所示）将家中的各种设备（如音视频设备、照明系统、窗帘控制、空调控制、安防系统、数字影院系统、影音服务器、影柜系统、网络家电等）连接到一起，提供家电控制、照明控制、通信远程管控、室内外遥控、防盗报警、环境监测、暖通控制、红外转发以及可编程定时控制等多种功能。对比传统住宅，智能家居在满足基础居住需求之外，整合建筑、通信、智能家电、设备自动化能力，实现全维度信息交互，还可优化能源消耗、降低用能开支。



图 1-5　手机＋APP 控制

（三）交通

自动驾驶是人工智能核心落地赛道。在现代交通体系下，无人车辆将逐步成为未来主流出行载体。车辆依托环境感知、交通标志识别、智能决策、路径规划等多重能力处置各类复杂路况，实现安全、高效自动驾驶（如图 1-6 所示）。



图 1-6　无人驾驶

（四）医疗

随着医疗技术的不断进步，人工智能在医疗领域的应用正发挥着重要的作用。首先，AI

可以用于医学影像诊断。通过深度学习算法，人工智能可以帮助医生更准确地诊断肿瘤、心脏病等疾病。其次，人工智能还可以用于疾病预测和风险评估。通过深度学习算法，人工智能医学影像工具可辅助医生更高效地筛查肿瘤、识别心脏病等病变，提升诊断参考准确度。最后，AI 在药物研发、手术机器人（如图 1-7 所示）和远程医疗等方面有着广泛的应用。



图 1-7　手术机器人

（五）教育

人工智能在教育领域中也有着广泛的应用。例如，人工智能可以通过分析学生的学习行为和知识点掌握情况，制定个性化的学习计划。这种计划可以基于学生的知识储备和学习进度，帮助学生更快地掌握知识点，提高学习效率。再如，人工智能还可以协助教师开展定制化的教学课程设计（如图 1-8 所示）。同时，人工智能可以协助教师梳理学生学情与学习诉求，完成定制化课程设计。此外，人工智能还可以辅助教师进行教学评估和学生成绩预测，为教师提供更为全面的教学支持。



图 1-8　人工智能教学课程设计

综上，伴随技术持续迭代，AI 落地场景将持续拓宽，不断深化，既能提升各行业生产效能，也能催生全新商业模式与就业岗位。

思考与练习

一、选择题

1. 在构建一个用于预测学生食堂菜品需求的智能推荐系统时，以下哪个选项最能体现“算力”作为“引擎”的核心作用？（　　）

A. 系统收集了学生过去一年的消费记录和口味偏好问卷。

B. 系统使用协同过滤算法来分析用户间的相似性并进行推荐。

C. 系统实时处理校园卡海量交易数据流，秒级完成模型推理。

D. 系统的预测结果被用来指导次日的食材采购计划。

2. 关于人工智能的“模拟法”，以下哪项描述是准确的？（　　）

A. 该方法通过人工编写详细的程序逻辑来实现特定功能，如国际象棋 AI。

B. 该方法的核心优势在于编程简单，易于调试和维护。

C. 该方法要求编程者具备生物学或认知科学的思考方式，去模拟智能的产生过程。

D. 该方法在处理简单、规则明确的任务时，比工程学方法更省力。

3. 一个 AI 系统被设计用来分析社交媒体帖子，不仅能判断其情感是“正面”还是“负面”，还能生成一篇针对该情感的回复文章。根据其功能划分，该系统最符合以下哪一类？（　　）

A. 仅决策式人工智能

B. 仅生成式人工智能

C. 同时具备决策式和生成式能力的人工智能

D. 弱人工智能

4. 与传统机器学习相比，深度学习最大的优势在于：（　　）

A. 需要的训练数据量更少。

B. 训练过程更快，对算力要求更低。

C. 能够自动从原始数据中学习和提取特征，无需人工特征工程。

D. 模型的可解释性更强。

5. 一个 AI 系统通过与环境的不断交互，根据获得的奖励（如得分）或惩罚（如失败）来调整其策略，以最大化长期累积奖励。这种学习方式被称为：（　　）

A. 监督学习　　B. 无监督学习　　C. 强化学习　　D. 迁移学习

二、思考题

1. 人工智能的算力为系统“引擎”、数据为系统“燃料”、算法为系统“大脑”。请结合“智能推荐系统”或“自动驾驶”等具体应用实例，详细论述这三大要素是如何协同工作，共同驱动人工智能系统实现其智能功能的。

2. 请分别说明决策式和生成式这两类 AI 的核心功能是什么，并各举一个本项目中提到的或生活中常见的应用例子，简要解释例子如何体现其功能。

项目二

探索视觉智能新境界：AI生成图像技术基本应用

项目导读

我们正身处一个由人工智能（AI）深刻塑造的视觉革命时代。从20世纪70年代艺术家哈罗德·科恩的AARON系统进行机械臂绘画，到如今仅凭一句文字描述就能生成媲美专业画师作品的AI生图技术，人工智能在视觉创作领域的飞跃令人惊叹。

本项目以“探索视觉智能新境界”为主题，旨在带领学生深入AI生成图像这一激动人心的技术前沿。我们将系统学习AI生图的发展脉络、核心技术原理与面临的挑战，并通过实践操作，掌握如何运用“提示词”这一关键指令与AI进行有效“对话”。通过本项目的学习，学生不仅能解锁AI强大的视觉创作能力，更能培养在人工智能生成内容的时代必备的创新思维、技术素养与批判性眼光，为未来的数字生活和创意工作做好准备。

学习指导

通过本项目的学习，你将：

- 理解AI生图的核心概念与发展脉络。
- 认识AI生图技术的创新性、多样性、高效性等特点，并能辨析其在数据、算力、伦理等方面面临的挑战。
- 能够独立使用“讯飞星火”“文心一言”等主流平台，完成从文本生成图像、以图生图的操作。
- 利用AI作为辅助工具探索无限的视觉表达可能性，同时，建立对AI生成内容的批判性认知。

任务一　AI 生成图像技术

一、AI 生图技术的发展

AI 生图技术的快速发展得益于深度学习、自然语言处理、计算机视觉等多个领域的突破性进展。学习者了解前沿动态，能够及时捕捉到这些领域的最新技术成果，并将其应用于文生图任务中，从而推动技术创新和进步。

（一）AARON 系统

AI 生图的历史可以追溯到 20 世纪 70 年代，艺术家哈罗德·科恩（Harold Cohen）研发 AARON 系统，该系统依靠算法生成画作指令，驱动机械臂完成绘画输出。AARON 系统的特点是使用机械臂和计算机程序进行创作，如图 2-1 所示。虽然早期的作品风格较为抽象，但它标志着 AI 绘画的初步尝试。

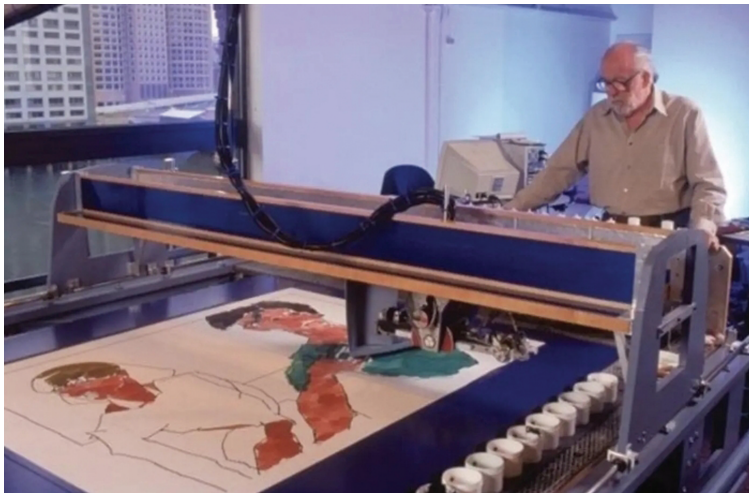


图 2-1　AARON 系统机械臂输出绘画

（二）基于深度神经网络的现代模型

随着时间的推移，AI 生图技术逐渐从基于规则的算法发展到基于深度神经网络的现代模型。2012 年，谷歌的吴恩达、Jeff Dean 等人采用无监督深度神经网络，在大量无标注的 YouTube 视频图像帧上进行训练，网络在训练过程中自发形成了对猫脸等高层语义特征高度敏感的神经元，其特征可视化结果如图 2-2 所示，即网络上常被提及的“模糊猫脸”。需要说明的是，该工作本质上属于特征学习与识别，而非图像生成，但它验证了深度神经网络能够从海量无标注数据中自动抽象出人类可理解的语义级特征，与同年在 ImageNet 视觉识别评测中取得突破的 AlexNet 一道，为此后 AI 生图技术的兴起奠定了重要基础，可视为该领域的早

期探索。

2015 年，谷歌推出了“深梦”（Deep Dream）图像生成工具（如图 2-3 所示），尽管它更像是一个高级滤镜，但它也标志着 AI 绘画技术的进步。

此后，AI 生图技术不断进步，出现了诸如 DALL-E、CLIP 和 Stable Diffusion 等先进的模型，这些模型能够根据输入的文本生成高度复杂和逼真的图像。



图 2-2　模糊猫脸的模型

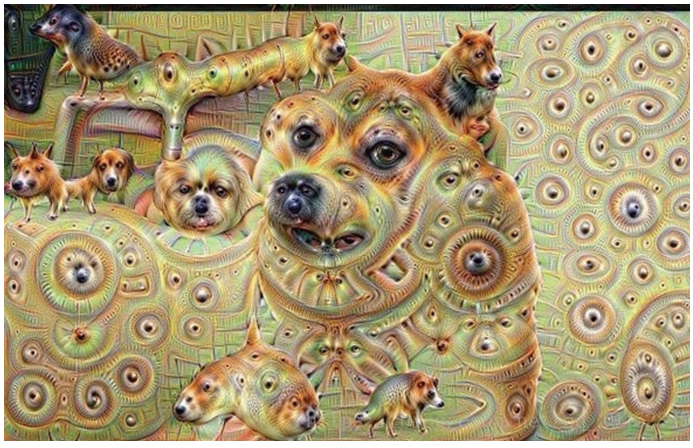


图 2-3　Deep Dream 生图的效果

（三）DALL-E 模型

2021 年，OpenAI 推出了 DALL-E 深度学习算法模型，能够直接从文本提示“按需创造”风格多样的图形设计，如图 2-4 所示的牛油果形状的椅子。DALL-E 的出现极大地扩展了 AI 生图技术的应用范围，使其在设计、广告、艺术创作等领域发挥了重要作用。

近年来，扩散模型成为 AI 生图技术重要发展方向。扩散模型的原理借鉴了物理学上的扩散现象，通过向训练图像中不断加入噪声，然后通过 AI 预测并逐步减少噪声，最终生成新的图像。这种技术不仅提升了图像生成的质量，也使得 AI 生图技术更加灵活。

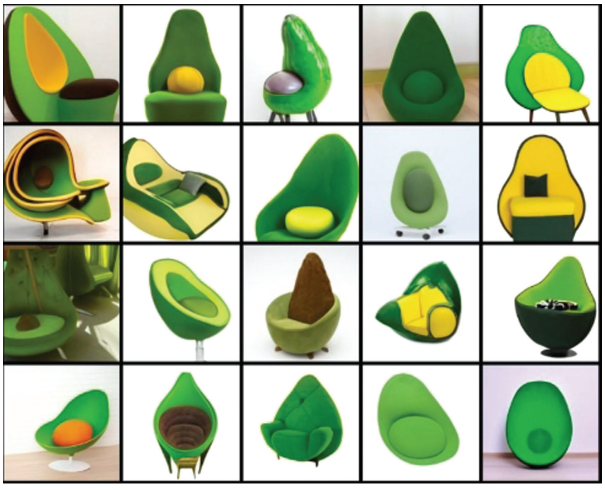


图 2-4　DALL-E 模型生图的效果

二、AI 生图的难点和挑战

随着人工智能生成内容（AIGC）领域的快速发展，特别是 AI 生图技术的进步，我们正

见证着一个前所未有的视觉创作新时代的到来。然而，在这项技术带来无限可能性的同时，也面临着诸多难题和挑战。

（一）数据质量问题

高质量、多样化的数据集是训练有效模型的基础，但获取这样的数据集往往成本高昂且耗时，不充分或偏差的数据可能导致生成图像存在偏见或者缺乏真实性。

（二）计算资源需求

当前先进的图像生成算法通常需要大量的 GPU 加速处理能力才能高效运行，高昂的硬件成本限制了个人用户及小型企业参与其中的机会，同时也增加了环境负担。

（三）创意控制难题

尽管 AI 模型已经能够在一定程度上理解文本描述并生成相应的图像，但要想完全实现文本与图像的完美对应，让用户直观地指定想要生成的具体风格或特征，仍然是一个巨大的挑战。缺乏灵活有效的输入方式可能会阻碍非专业人士使用该技术进行创作。此外，模型在生成图像时也容易陷入局部最优解，导致生成的图像缺乏新颖性和惊喜感。

（四）伦理道德考量

随着 AI 生图技术的普及和应用，版权和伦理问题也日益凸显。滥用 AI 生图技术可能引发版权争议、隐私泄露等问题，必须建立健全相关法律法规来规范其应用范围，保护创作者权益和个人信息安全。因此，在推动 AI 生图技术发展的同时，我们也需要关注并解决这些问题。

（五）技术门槛与资源消耗

第一，AI 生图底层技术横跨深度学习、自然语言处理、计算机视觉等多学科，技术门槛较高；第二，前期算力、数据集研发成本高昂，压缩了个人、小型团队的参与空间。

三、AI 生图特点

AI 生图利用深度学习算法和神经网络等技术，通过已有的数据集学习和理解绘画的特征、风格和规律，从而创作出独特的艺术作品。AI 绘图技术以其创新性、多样性、高效性等特点，正在成为艺术创作领域的重要工具。

（一）创新性

AI 绘图技术能够突破传统艺术的限制，创造出前所未有的视觉效果和风格。这种创新性体现在以下几个方面：

独特的视觉表达：AI 可以生成具有独特纹理、色彩和形状的图像，这些图像在传统艺术中很难实现或从未出现过。

融合多种元素：AI 可以将看似不相关的元素融合在一起，创造出全新的艺术形式，如将自然景观与科幻元素结合。

实验性创作：艺术家可以利用 AI 进行实验性创作，探索未知的艺术领域，打破常规的创

作模式。

（二）多样性

AI 绘图可以模拟多种艺术风格，从古典到现代，从抽象到具象，几乎无所不能。这种多样性为用户提供了广泛的选择。

风格广泛：无论是印象派、立体主义还是超现实主义，AI 都能模仿并生成相应的艺术作品。

文化融合：AI 可以融合不同文化的艺术元素，创造出跨文化的艺术作品，促进文化交流和理解。

个性化定制：用户可以根据自己的喜好选择特定的艺术风格，甚至创造属于自己的独特风格。

（三）高效性

AI 绘图工具可以迅速将文本描述转换为图像，极大地提高了创作效率。

快速迭代：艺术家可以快速生成多个版本的图像，从中选择最满意的一个，节省了大量的时间和精力。

即时反馈：AI 绘图工具可以提供即时的视觉反馈，帮助艺术家更好地理解和调整他们的创意。

批量生成：对于需要大量图像的项目，如广告或社交媒体内容，AI 可以快速生成多幅图像，满足大规模生产的需求。

（四）可定制性

用户可以通过输入特定的文本提示或参数来定制 AI 生成的图像，满足个性化需求。

细节控制：用户可以通过指定颜色、形状、大小等细节来精确控制生成的图像。

情感表达：通过描述情感或氛围（如“忧郁的”或“欢快的”），用户可以引导 AI 生成具有特定情感色彩的图像。

风格混合：用户可以尝试将不同的艺术风格混合在一起，创造出独特的视觉效果。

（五）交互性

一些 AI 绘图工具支持实时预览和编辑功能，用户可以即时调整和优化生成的图像。这种交互性提高了用户的参与度和满意度。

实时预览：用户可以在生成过程中实时查看图像的变化，及时做出调整。

编辑和修改：生成的图像可以轻松地进行编辑和修改，以符合用户的期望。

用户反馈循环：用户的反馈可以被用来改进 AI 模型，使其更好地理解和满足用户需求。

（六）跨平台性

许多 AI 绘图工具支持云端服务，可以在不同设备和平台上使用，方便用户随时随地进行创作。这种跨平台性为此提供了极大的便利。

多设备访问：用户可以在电脑、平板或手机上使用 AI 绘图工具，不受地点限制。

数据同步：云端服务允许用户在不同设备之间同步他们的项目和进度，确保工作的连续性。

协作创作：跨平台的特性也使得多人协作变得更加容易，艺术家们可以远程合作，共同完成项目。

（七）学习性

AI 绘图技术不断学习和进化，随着时间的推移，其生成的图像质量和创意水平会不断提高。

持续改进：AI 模型可以通过分析大量的艺术作品和用户反馈来不断优化其算法。

适应新趋势：AI 可以快速适应新的艺术趋势和风格，保持创作的现代性和相关性。

个性化学习：一些 AI 系统可以根据用户的偏好和使用习惯进行个性化学习，提供更加定制化的服务。

（八）辅助性

AI 绘图不仅可以独立创作艺术作品，还可以作为艺术家的辅助工具，帮助他们探索新的创意和灵感。

四、AI 生图提示词

AI 生图提示词也被称为 AI 绘画指令或提示短语，是用户向 AI 生图工具或软件提供的具体指令，用以指导 AI 生成符合用户需求的图像。这些提示词对于 AI 来说至关重要，因为它们直接决定了生成的图像的风格、内容、色彩、构图等多个方面。通过合理运用这些规则，可以提高生成内容的质量和相关性。

（一）提示词的语法规则

提示词的语法规则包括基本语法规则、权重调整以及特殊语法。提示词通常由内容型提示词和标准化提示词构成，它们共同作用以引导 AI 生成期望的内容。下面将具体介绍提示词的语法规则：

1. 基本语法规则

● 自然语言写法：使用完整的句子来描述想要生成的内容，例如“一个穿蓝色裙子的女孩在草地上玩耍”。

● 标签语言写法：通过关键词或短语来指定内容，如“1 个女孩，蓝色连衣裙，草地，玩耍”。

● 权重写法：通过小括号（）、中括号［］或数字来调整提示词的权重，如“1 个女孩（红色连衣裙：1.2），草地，玩耍”。

● 渐变写法：用于在生成过程中逐渐改变某些特征，“1 个女孩，草，玩耍，［红袍：黑甲：0.1］”。

● 交替写法：用于融合不同的元素，如“［狗 | 猫］，草，玩耍”。

● 组合写法：直接融合多个提示词的结果，并支持权重调整，“（狗：1.2）和猫，草，玩耍”。