人工智能通识与 AIGC 应用

王　磊　何宗要　主　编
王德如　薛明志　马　骏　副主编

清華大學出版社
北　京

内 容 简 介

本书第Ⅰ篇主要涵盖人工智能概述、伦理、大语言模型、智能体、生成式AI等内容，每个项目以实际场景切入，引导学生分析问题并实践。第Ⅱ篇聚焦文本生成，数据分析，图像、语音、视频生成，AI辅助编程等实践技能，帮助学生使用低代码工具(如DeepSeek、Midjourney)解决实际问题。第Ⅲ篇适合信息类专业的学生深入学习。本书案例均源于校园生活与产业需求，有助于强化学生将技术与场景相结合的能力。项目1~11使用零代码和低代码工具完成，降低了非信息技术专业学生的学习难度。本书免费提供教学资源，可扫封底二维码下载。

本书可以作为高等职业院校、应用型本科、中等职业学校人工智能通识课的教材。

图书在版编目(CIP)数据

人工智能通识与AIGC应用 / 王磊, 何宗要主编.

北京 : 清华大学出版社, 2025. 9. -- ISBN 978-7-302-70276-4

Ⅰ. TP18

中国国家版本馆CIP数据核字第202531GF04号

责任编辑：王　军

封面设计：孔祥峰

版式设计：思创景点

责任校对：成凤进

责任印制：宋　林

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：大厂回族自治县彩虹印刷有限公司

经　　销：全国新华书店

开　　本：190mm×260mm　　**印　　张：**11.25　　**字　　数：**267千字

版　　次：2025年9月第1版　　**印　　次：**2025年9月第1次印刷

定　　价：49.80元

产品编号：113384-01

指导委员会

主　任

蒋笃运

副主任

孔凡士　马广建

委　员

厉　励　苏新留　张建伟　刘荣宁　李　森　李立峰
谷建光　王广国　陶保富　王泽民　刘珊珊

编写委员会

主　编

王　磊　何宗要

副主编

王德如　薛明志　马　骏

编　委

李纪云　张　巍　冯明卿　孙建国　张新成　赵　恒
付晓炎　蒋清健　郭静博　王磊杰　简艳英　胡仁喜
李　冰　张馨怡　杨　雨　齐港谋　李辉利　刘　悦
李　雪　张成林　张沛杰　谭　昱

序 言 PREFACE

人工智能通识教育：赋能高技能人才的新时代序章

我们正身处一个被智能技术深刻重塑的时代。人工智能 (AI) 已从实验室走向千行百业，从尖端概念演变为驱动产业升级的核心引擎。从车间智能机器人的精准操控到供应链预测模型的优化决策，从智慧医疗影像的辅助诊断到客户服务过程中自然语言处理的流畅交互，人工智能正以前所未有的广度和深度融入职业场景，重新定义工作内涵与技能图谱。

面对这场深刻的技术与产业革命，职业教育肩负着前所未有的使命：培养具备人工智能素养的新时代高技能人才。这些人不仅是熟练的操作者，更应是理解智能技术逻辑、善用智能工具解决实际问题，并能对其应用边界与社会影响保持清醒认知的复合型人才。这正是高职院校开设人工智能通识课的根本逻辑与时代必然。

本书的编写严格遵循《人工智能通识课程体系规范》的团体标准，紧密对接《职业院校人工智能应用指引》的教育教学改革，并深度回应技术快速迭代与行业深度融合的双重挑战。我们深知，人工智能通识教育绝非简单的“科普”，是赋能个体、塑造未来的关键一步。它关乎每一位高职学生能否在智能时代赢得职业尊严、实现人生价值，也关乎我国能否从“制造大国”迈向“智造强国”，培养出规模宏大、素质优良的高技能人才大军。本书的核心目标在于：

(1) 奠基认知，破除神秘：系统、准确地阐释人工智能的基本概念与发展脉络、核心技术、典型应用，帮助学生构建清晰、科学的技术认知框架，破除对“黑箱”的盲目崇拜或恐惧。

(2) 聚焦应用，赋能实践：紧密结合高职学生未来职业领域，结合校园生活场景，引导学生思考如何优化流程、提升效率、创造价值。

(3) 塑造思维，提升素养：着力培养“人工智能思维”，包括理解数据驱动决策的价值、掌握问题拆解与模型化思路，以及利用 AI 工具探索创新的意识。

(4) 强化伦理，明确责任：深刻探讨人工智能应用引发的伦理挑战(如算法偏见、隐私保护、安全可控性等)与社会影响，引导学生树立负责任的技术价值观，理解在职业活动中应用AI的道德边界与社会责任，成为技术向善的践行者。

(5) 启迪终身学习，应对未来变革：激发学生对人工智能领域持续探索的兴趣，培养其适应技术快速迭代的自主学习能力与开放心态，为职业生涯的可持续发展奠定坚实基础。

为实现上述目标，本书在内容设计与编排上力求体现以下几点。

- 专业性：内容科学、严谨，概念表述精准，技术原理阐释深入浅出，确保知识体系的准确性与前沿性。紧密跟踪大模型、AIGC 等技术的最新发展趋势及其在产业中的应用潜力。
- 普适性：立足高职学生知识背景，避免艰深的数学推导，语言力求清晰、生动、易懂，强调直观理解与实际联系。大量采用来自校园的案例，确保不同专业背景的学生都能学懂、会用、有共鸣。
- 引领性：超越单纯的知识传授，强调思维塑造与能力培养。突出职业教育的类型特色，强化 AI 作为通用赋能技术在提升职业综合能力方面的核心作用。
- 实践导向：设计丰富的课堂活动、讨论议题和基于真实场景的实训项目，强调学以致用，在做中学，在思中悟。

本书在河南省人工智能学会的统筹下，组织相关领导、专家团队完成，凝聚了高职院校一线教师、人工智能领域的专家和学者，以及行业企业技术骨干的集体智慧。我们深入企业调研，了解真实岗位对 AI 素养的需求；我们反复研讨课程大纲，确保内容精准匹配高职人才培养定位与通识课教学目标；我们精心打磨每一个案例与表述，力求在专业性与可读性之间找到最佳平衡点。

本书的编写过程亦是对人工智能教育本质的持续叩问：在技术狂飙突进的时代，职业教育如何坚守育人为本的初心？我们的答案是：不仅要教会学生如何与机器“对话”，更要启迪他们思考如何让技术服务于人的全面发展与增进社会福祉；不仅要培养学生驾驭工具的能力，更要塑造其作为技术主体的责任与远见。 我们期望，通过这门课程的学习，学生能够做到以下几点。

- 看得懂：洞悉身边 AI 应用的底层逻辑，不再雾里看花。
- 用得上：掌握运用 AI 工具提升专业效率、解决工作难题的基本方法。
- 想得深：对技术发展保持独立思辨，明晰应用的伦理边界与社会责任。
- 走得远：奠定在智能化浪潮中持续学习、适应变革、创新发展的能力基石。

本书是我们对这一时代命题的探索与实践。我们深知，人工智能领域日新月异，通识教育亦须与时俱进。我们热忱欢迎广大师生、行业同仁在使用过程中提出宝贵意见，共同推动人工智能通识教育在高职沃土上生根发芽、枝繁叶茂，为培养堪当民族复兴重任的大国工匠、能工巧匠贡献智慧与力量！

是为序。

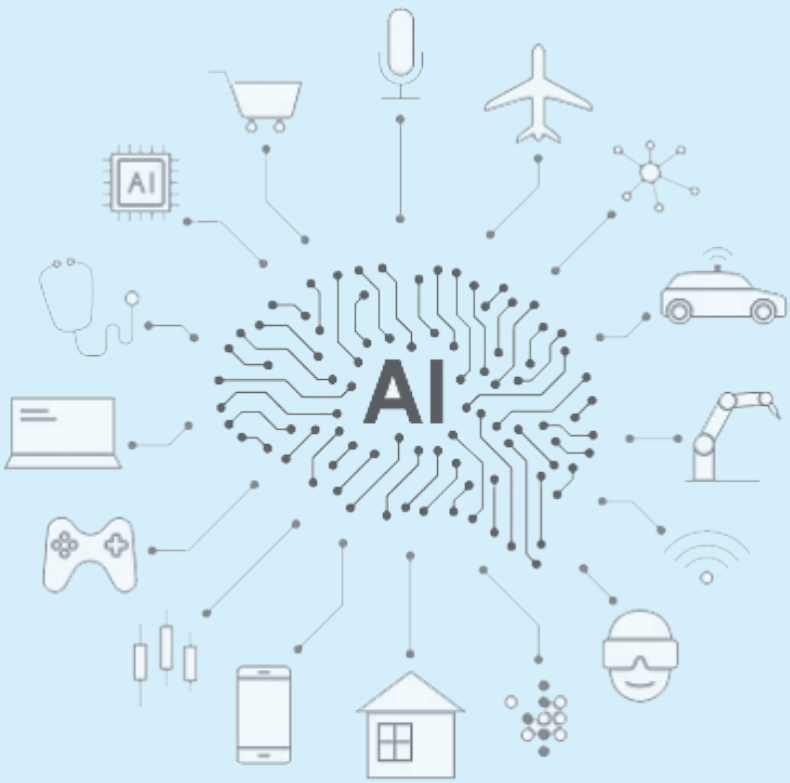
教材编写委员会

2025 年 7 月 16 日

目 录 CONTENTS

第 | 篇

基础知识



项目 1
人工智能概述

案例导学 食堂的“选择困难症”

场景描述

某高职食堂日均接待3000人，但面临以下问题

- 菜品浪费：40%的学生因不合口味倒掉部分饭菜
- 排队拥堵：午餐高峰排队超20分钟
- 营养失衡：学生偏爱油炸食品，绿叶菜剩余量大

解决思维

问题：食堂浪费 → 数据：消费记录 → 算法：推荐系统 → 工具：Python、Pandas → 应用：手机推送

解决方案

部署AI智能推荐系统，根据学生口味、消费记录、营养数据，实现以下功能

- 个性化菜品推荐(手机端实时推送)
- 动态排队预测(引导分流至人少窗口)
- 营养均衡分析(周报推送饮食建议)

个性化菜品推荐：系统会根据学生的口味偏好和消费记录，推荐符合个人口味且营养均衡的菜品，并通过手机端实时推送，帮助学生选择更合适的菜品。

动态排队预测：系统能够实时监测食堂各窗口的排队情况，并引导学生分流至人少的窗口，从而减少排队时间，提高就餐效率。

营养均衡分析：系统会分析学生一周的饮食数据，生成营养报告，并通过周报的形式推送给学生，帮助他们调整饮食习惯，促进健康饮食。

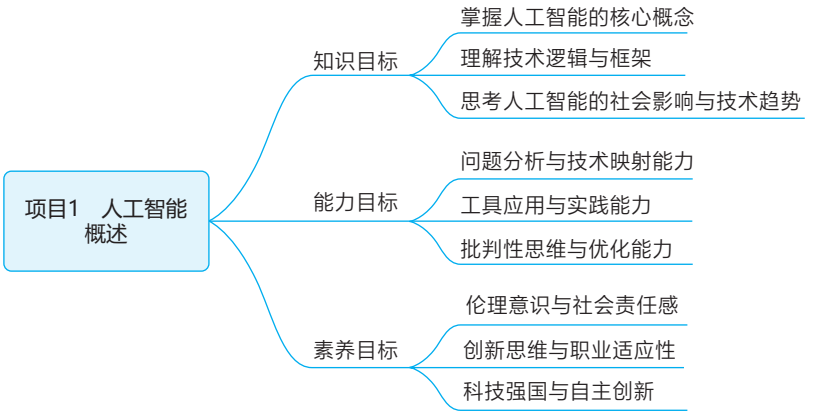
学生讨论

1. AI 是如何获取和分析学生口味偏好、消费记录等数据的？数据的准确性和完整性如何保证？

2. 如果推荐系统不够准确，应该如何优化？如何根据学生的反馈进一步调整推荐算法？

3. 收集学生数据时，如何确保数据隐私和安全？如何平衡数据利用与学生隐私保护之间的关系？

学习导图



知识学习

1.1　人工智能的发展

1.1.1　人工智能的起源与早期思想

人工智能(artificial intelligence，AI)作为一个综合性的科学领域，其思想萌芽和初步探索主要集中在20世纪中叶。这一时期的关键进展为人工智能的诞生奠定了重要的理论和概念基础。

1. 麦卡洛克-皮茨神经元模型

“人工智能”概念被正式提出之前，学术界已有相关的理论探索。1943年，神经生理学家沃伦·麦卡洛克(Warren McCulloch，见图1-1)和逻辑学家沃尔特·皮茨(Walter Pitts，见图1-2)发表了论文“神经活动中思想的逻辑演算”，提出了第一个人工神经元的可计算模型，即麦卡洛克-皮茨神经元模型。他们证明了这种由简单二元(开和关)单元组成的网络能够完成任何可计算的逻辑功能。这项工作为后来的连接主义(神经网络)研究提供了重要的早期理论启发。

图 1-1　沃伦·麦卡洛克

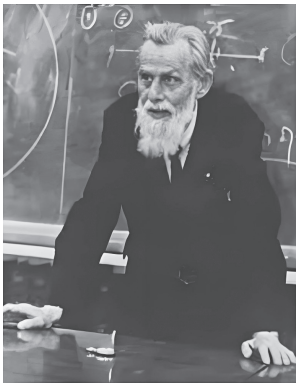


图 1-2　沃尔特·皮茨

2. 图灵测试

英国数学家、计算机科学的先驱艾伦·图灵(Alan Turing，见图1-3)在其1950年发表的论文“计算机器与智能”中，提出了一个判断机器是否具有与人类无法区分的智能行为的思想实验，即著名的图灵测试(Turing Test)。图灵设想，如果一台机器能够通过文本界面与人类进行对话，且在一定时间内，人类测试者无法可靠地分辨出与其交流的是人还是机器，那么就可以认为这台机器具备了智能。图灵测试并非仅定义智能本身，而是为衡量机器智能提供了一个可操作的、行为主义的标准，对人工智能的哲学思辨和长远发展产生了深远影响，为人工智能的发展奠定了基础。

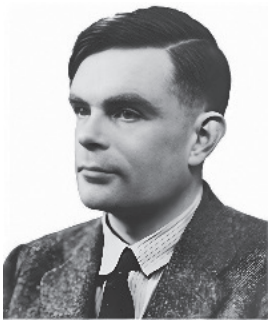


图 1-3　图灵

3. 达特茅斯会议

1956年，约翰·麦卡锡(John McCarthy)等人在“达特茅斯夏季人工智能研究计划”会议(以下简称达特茅斯会议)中首次提出了“人工智能”这一术语，并将其定义为“制造智能机器，特别是智能计算机程序的科学与工程”。这标志着“人工智能”作为一个特定研究领域的名称被确立下来。会议聚集了一批计算机科学家、数学家和逻辑学家，他们共同探讨了用机器模拟人类学习、语言理解、问题求解、模式识别、创造力等各种智能行为的可能性与实现途径。会议的核心信念如下：学习的每一个方面或智能的任何其他特征原则上都可以被精确地描述，从而可以用机器来模拟它。达特茅斯会议(参会者见图1-4)被广泛视为人工智能诞生的标志，为人工智能后续数十年的探索与发展奠定了组织基础和智力框架，开启了人工智能研究的新纪元。

图 1-4　诞生人工智能的达特茅斯会议参会者 (注：图由大模型生成)

1.1.2　人工智能的发展历程与关键突破

自人工智能学科诞生以来，其发展并非一帆风顺，而是呈现波浪式前进、螺旋式上升的特点，经历了多次高潮与低谷的交替。理解这一历程中的关键时期、重要突破以及所面临的挑战，对于把握当前人工智能的发展态势至关重要。

1. 黄金时期(20世纪50年代末至70年代初)

达特茅斯会议之后的十多年，迎来了人工智能的黄金时期，在这一时期，人工智能取得了许多重要突破。

1956年，赫伯特·西蒙(Herbert Simon)和艾伦·纽厄尔(Allen Newell)共同开发了“逻辑理论家”(Logic Theorist)。这是第一个能够证明数学定理的计算机程序，它证明了《数学原理》第二章中的38条定理，其中一条甚至找到了比罗素原文更简洁的证明步骤，标志着机器的符号推理能力得到初步展现。

1959年，纽厄尔和西蒙等人开发了通用问题求解器 (General Problem Solver，GPS)，试图模拟人类解决问题的通用方法(如手段-目的分析)，这是早期问题求解研究的重要里程碑。

1958年，约翰·麦卡锡开发了LISP语言，该语言具有强大的符号处理能力和灵活的列表结构，成为人工智能研究的重要工具，也成为之后几十年人工智能领域主要的编程语言，在人工智能语言方面取得了重大进展。

此外，早期的专家系统也开始出现，如1968年斯坦福大学的爱德华·费根鲍姆(Edward Feigenbaum)领导的研究小组研发完成的专家系统DENDRAL，该系统用于化学分子结构分析，为以后专家系统的开发树立了榜样。

2. 低谷期(20世纪70年代初至80年代初)

早期的成功使得人们对人工智能的期望过高，随着研究的深入，一系列难以逾越的障碍开始显现，导致人工智能的发展进入了低谷期，即所谓的“AI寒冬”。

当时的计算机性能有限，无法处理复杂的智能任务。许多AI问题(如搜索、规划)随着问题规模的增大，其计算复杂度呈指数级增长，使得实际求解变得非常困难。同时，人工智能的研究也面临着一些理论和技术难题，如知识表示、推理机制等。这一时期，人工智能的研究进展缓慢，资金投入减少，研究热情降低。

3. 知识工程与专家系统的兴起(20世纪80年代)

经历了“AI寒冬”，人工智能研究者将目标转向更具体、更可控的领域。随着计算机硬件性能的提升和知识工程方法论的成熟，专家系统(见图1-5)迎来了发展的黄金时期。

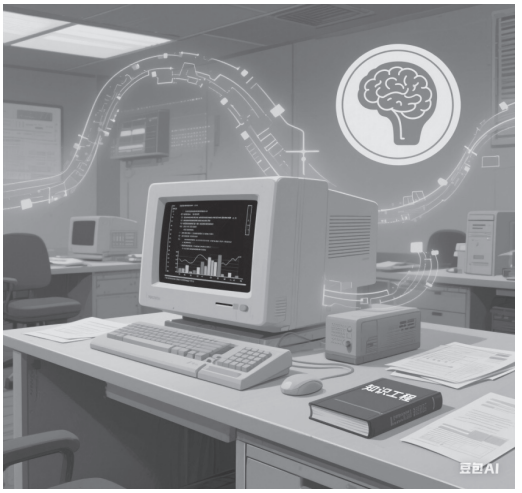


图 1-5　专家系统示意图 (注：图由大模型生成)

20世纪80年代，随着计算机性能的提高和知识工程的发展，人工智能迎来了新的发展机遇。专家系统成为这一时期的主要研究方向，它通过模拟专家的知识和经验，解决特定领域的复杂问题。

MYCIN(始于20世纪70年代初，80年代应用趋于成熟)由斯坦福大学的爱德华·肖特利夫(Edward Shortliffe)等人领导开发，是一个用于诊断细菌感染并推荐合适的抗生素治疗方案的专家系统。它通过运用包含约600条规则的知识库，模拟传染病专家的诊断逻辑和决策过程，在特定医疗领域展现了与人类专家相当的性能，并引入了“置信因子”来处理具有不确定性的信息，是早期医学专家系统研究的典范。

XCON(原名R1，始于20世纪70年代末，在80年代广泛部署)由卡内基梅隆大学的约翰·麦克德莫特(John McDermott)为数字设备公司(Digital Equipment Corporation，DEC)开发，是一个用于自动配置VAX系列小型计算机订单的专家系统。它能够根据客户需求选择合适的组件并确保它们之间的兼容性，极大地提高了配置的准确性和效率，每年为DEC公司节省了数千万美元的成本，是专家系统在工业界成功商业化应用并产生巨大经济效益的标志性案例。

4. 机器学习的兴起(20世纪90年代至21世纪第一个十年)

虽然专家系统取得了成功，但其知识获取瓶颈(依赖人工构建知识库)和领域局限性也逐渐暴露。与此同时，连接主义(神经网络)在经历了早期的沉寂后开始复兴，并与统计学习方法相结合，推动机器学习成为人工智能的主流研究方向之一。机器学习的目标是让计算机通过数据学习规律，从而实现智能决策。

1997年，IBM的“深蓝”(Deep Blue)计算机战胜了国际象棋世界冠军卡斯帕罗夫(Garry Kasparov)，展示了机器学习在复杂任务中的强大能力(见图1-6)。此外，支持向量机(SVM)、决策树等机器学习算法也在这一时期得到了广泛应用。

图 1-6　“深蓝”战胜棋王卡斯帕罗夫 (注：图由大模型生成)

5. 深度学习的突破(21世纪第二个十年至今)

21世纪第二个十年以来，基于大规模数据、强大算力(特别是GPU的应用)以及算法的改进，深度学习取得了革命性的突破，将人工智能推向了前所未有的新高度。

2012年，深度学习在图像识别领域取得了重大突破，Geoffrey Hinton团队的AlexNet在ImageNet竞赛中取得了优异成绩，开启了深度学习的热潮。深度学习通过构建多层神经网络，能够自动学习数据的特征表示，从而执行更高效的智能任务。

近年来，深度学习不仅在传统的图像识别、自然语言处理、语音识别等核心AI领域取得了里程碑式的成果，还在科学发现(如AlphaFold预测蛋白质结构)、药物研发、材料科学、气候变化建模等众多交叉学科中展现出巨大的应用潜力，持续推动着人工智能技术的边界拓展和产业的快速发展。

1.1.3　人工智能的未来展望与挑战

人工智能作为引领新一轮科技革命和产业变革的战略性技术，其未来发展充满了无限可能，也伴随着诸多需要审慎应对的挑战。

1. 技术融合与创新

未来，人工智能将与其他新兴技术深度融合，如物联网(IoT)、大数据、区块链等，形成更强大的智能系统。例如，通过物联网设备收集的数据，结合人工智能算法，可以实现智能家居、智能交通等应用场景的智能化管理。

2. 应用领域的拓展

人工智能的应用领域将不断拓展，除了现有的医疗、教育、金融等领域，还将在农业、环保、能源等更多领域发挥重要作用。例如，采用人工智能技术可以实现精准农业，提高农作物产量；可以利用人工智能进行环境监测和污染治理，保护生态环境。

3. 伦理与社会影响

随着人工智能的广泛应用，其伦理和社会影响也日益受到关注。例如，人工智能可能导

致就业结构的变化，一些传统职业可能会被取代。同时，人工智能系统的决策过程可能涉及隐私保护、公平性等问题，因此，未来需要加强对人工智能伦理的研究和规范，确保人工智能技术的健康发展。

1.2　人工智能的定义与分类

1.2.1　人工智能的多视角定义

人工智能是一个多维度、跨学科的概念，其定义因研究领域和应用场景的不同而有所差异。以下是几种常见的定义。

1. 从智能行为的角度

人工智能可以被定义为使计算机系统能够执行通常需要人类智能的任务。这些任务包括但不限于语言理解、视觉感知、学习、推理、规划、决策和创造性思维。例如，语音助手(如Siri)能够理解自然语言并做出回应，推荐系统(如抖音)能够根据用户行为预测其兴趣并推荐内容，这些都是人工智能在实际应用中的体现。

2. 从功能模拟的角度

人工智能是计算机科学的一个分支，它研究如何使计算机模拟人类的智能行为和思维过程。该定义强调了人工智能的目标是通过计算机技术实现类似人类的智能功能，例如通过机器学习算法让计算机自动学习和优化，或者通过深度学习模型模拟人类大脑的神经网络结构。

3. 从应用领域的角度

人工智能也可以被定义为一系列技术的集合，这些技术能够解决复杂的、需要智能决策的问题。例如，在医疗领域，人工智能用于疾病诊断和治疗方案推荐；在交通领域，人工智能用于智能交通管理和自动驾驶；在金融领域，人工智能用于风险评估和投资决策。这些应用展示了人工智能在不同领域的广泛影响力。

1.2.2　人工智能的分类

为了更清晰地理解人工智能的当前能力、发展阶段以及未来潜能，可以从不同的维度对其进行分类。根据人工智能系统所能达到的智能水平或其能力的广度，人工智能可分为以下几类。

1. 弱人工智能

弱人工智能(Narrow AI)也被称为专用人工智能或应用型人工智能，是指专注于执行特定、预定义任务或在有限领域内解决特定问题的人工智能系统。这些系统被设计和用于完成一项或少数几项高度专门化的工作，并在这些特定任务上达到甚至超越人类的水平。然而，它们缺乏真正的意识、自我认知能力以及跨领域的通用学习和推理能力，不具备人类的普适性智能。目前我们所能接触到的和被广泛应用的所有人工智能系统，都属于弱人工智能的范

畴，如语音助手、图像识别系统、推荐系统等。

2. 强人工智能

强人工智能(Artificial General Intelligence，AGI)又称通用人工智能，是指具备与人类相当的、全面的认知能力的人工智能系统。AGI系统将能够理解、学习和应用知识，完成人类能够执行的智力任务，具备自主学习、抽象思考、常识推理、计划、沟通乃至创造等多种通用智能特征。它不再局限于特定领域，而是能够像人类一样灵活地适应各种未知或全新的情境。目前，强人工智能仍处于研究阶段，尚未实现商业化应用。

3. 超人工智能

超人工智能(Artificial Super Intelligence，ASI)是指在几乎所有领域都比人类更聪明、更高效的智能系统。这种智能不仅在速度和效率上超越人类，更可能在智慧、创造力、解决复杂问题及战略规划等多个维度达到人类无法企及的水平。这种系统目前还处于科幻和理论探讨阶段，但引发了广泛的关注和讨论，特别是在AI安全、AI对齐(确保AI目标与人类价值观一致)以及未来社会治理等议题上。

1.3　人工智能的核心要素

人工智能系统的构建与实现依赖以下四大相互关联、不可或缺的核心要素。这些要素共同构成了人工智能技术发展的基础。

1. 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(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水平。例如，在食堂的智能推荐系统中，其有效运行依赖大量高质量的数据。这些数据包括但不限于学生的历史消费记录(菜品、时间、频率)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(见图1-7)。这些数据是系统进行个性化推荐、预测菜品需求以优化采购与备餐计划的基本依据。

图 1-7　某食堂数据可视化系统数据展示界面 (注：图由大模型生成)

2. 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，系统需要实时处理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(见图1-8)，这需要相当强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 1-8　停车场车位识别摄像头采集数据 (注：图由大模型生成)

3. 算法

算法是人工智能的“大脑”，是实现智能决策的核心逻辑。算法定义了机器从数据中学习、推理和决策的逻辑与方法。算法通过对数据进行学习，提取特征、发现模式与规律，从而实现预测、分类、聚类、规划、优化等特定功能。例如，图书馆的座位管理系统的核心是一套座位分配与调度算法，该算法通过实时监测各座位传感器的使用数据，结合学生的预约信息、历史使用模式以及不同区域座位的热度等因素，动态地优化座位分配策略(见图1-9)。该系统可以引导学生至空闲座位，处理预约请求，并可能对长时间占用但未实际使用的座位进行管理。这种基于算法的智能调度是提升资源利用效率和用户服务体验的关键技术手段。

图 1-9　图书馆实时座位监测 (注：图由大模型生成)

4. 场景

场景是人工智能赖以行动和体现价值的“身体”，它构成了人工智能应用的具体环境，并明确了其需要满足的特定需求。场景不仅定义了人工智能系统需要解决的特定问题、服务的用户群体以及期望达成的应用目标，也规定了人工智能“身体”活动的边界和条件。人工智能系统的设计和优化需要紧密结合实际应用场景，以满足特定的需求。例如，在校园食堂场景中，智能推荐系统的构建需要深入分析学生群体的饮食偏好(如口味、菜系)、营养需求(如低脂、高蛋白)、消费水平以及高峰期快速点餐等行为特性(见图1-10)。

图 1-10　校园食堂智能推荐 (注：图由大模型生成)

1.4　人工智能的应用

人工智能作为一项极具影响力的前沿技术，已广泛且深入地融入人们的生活和工作，为人们带来了极大便利，推动了行业的创新发展，展现出了巨大的应用价值。人工智能在不同行业的典型应用场景参见表1-1。

表 1-1　人工智能在不同行业的典型应用场景

行业领域	涉及技术	应用场景
智能制造	工业机器人、计算机视觉、预测性维护	自动化生产线、缺陷检测、设备故障预测
智慧交通	自动驾驶、智能交通管理、智慧停车	无人驾驶汽车、实时路况优化、停车引导
智慧金融	风险评估、欺诈检测、个性化推荐	贷款审批、反洗钱系统、投资建议
智慧教育	智能辅导系统、在线答疑、学情分析	AI助教、作业批改、学习路径推荐
智慧家居	语音助手、智能家居联动、家庭安防	智能音箱、自动调节灯光、监控报警
智慧零售	商品推荐、库存管理、无人超市	个性化推荐、智能补货、自助结账
智能客服	自动回复、情感分析、多语言支持	在线聊天机器人、电话客服、常见问题解答
无人驾驶	环境感知、路径规划、紧急应对	特斯拉Autopilot、滴滴无人车
智慧物流	路径优化、仓储管理、无人机配送	京东无人仓、顺丰无人机快递
智慧安防	视频监控、身份验证、入侵检测	实时行为识别、人脸识别、网络攻击预警

续表

行业领域	涉及技术	应用场景
智慧医疗	影像诊断、药物研发、健康管理	癌症筛查、新药开发、远程诊疗
智慧环保	污染监测、资源回收、生态保护	空气质量分析、垃圾分类、野生动物保护
智慧旅游	个性化推荐、智能导览、人流预测	景点推荐、虚拟导游、景区流量控制
智慧农业	精准农业、病虫害防治、产量预测	无人机喷洒农药、土壤监测、作物生长监测

实训任务

任务 1　了解生活领域的 AI 应用

目前，人工智能已融入人们生活的方方面面，给我们带来很多便利与乐趣，以下列举几个典型的应用场景。

1. 自动驾驶

AI在自动驾驶领域的典型应用以特斯拉Autopilot为代表，车辆通过摄像头、毫米波雷达与超声波传感器的深度融合，实时感知车道线、交通标志及周边障碍物(见图1-11)。例如在拥堵的晚高峰路段，系统不仅能自动与前车保持安全车距，还会根据驾驶者日常偏好的激进或保守风格，动态调整变道频率和加速度。这种基于环境数据与用户习惯的双重优化，既降低了疲劳驾驶风险，又逐步培养了人车互信的协同关系。

图 1-11　自动驾驶示意 (注：图由大模型生成)

2. 语音助手

AI在语音助手领域的典型应用以Siri为代表，人们只需要通过简单的语音指令，就能完

成诸如查询天气、设置提醒、拨打电话等操作。例如，早晨起床时，用户无须手动操作手机，可以直接对 Siri 说 “查询今天的天气”，Siri 便会快速给出当地的天气信息，让用户提前做好出行准备(见图1-12)。这一应用极大地提升了人们获取信息和操作设备的便捷性，使生活更加高效。

图 1-12　手机语音助手 Siri(注：图由大模型生成)

3. 推荐系统

抖音的推荐系统依据用户的浏览历史、点赞、评论等数据，精准分析用户的兴趣偏好，为用户推送个性化的视频内容。用户在抖音上浏览美食视频较多时，推荐系统会不断推送各类美食制作、探店等相关视频，满足用户的兴趣需求，增加用户对平台的黏性。

4. 人脸识别

支付宝的人脸识别技术用于支付验证环节，用户进行支付操作时，无须输入烦琐的密码，只需要刷脸即可完成支付(见图1-13)。这种支付方式不仅快捷，还提高了支付的安全性，减少了密码泄露带来的风险。

图 1-13　人脸识别完成支付 (注：图由大模型生成)

任务 2　了解行业领域的 AI 应用

1. 工业质检

在制造业领域，人工智能技术能够对生产线上的产品进行快速、准确的质量检测。例如，利用计算机视觉和深度学习算法，识别产品表面的缺陷、尺寸偏差等问题。相比传统的人工质检方式，人工智能质检效率更高，检测精度更稳定，能够及时发现产品质量问题，降低次品率，提高企业的生产效益。

2. 医疗影像

在医疗领域，人工智能可辅助医生对医学影像进行分析。例如，通过对 X 光、CT、MRI 等影像的识别，帮助医生更准确地检测疾病，如肺部的微小肿瘤、脑部的病变等。这有助于提高疾病的诊断准确率，为患者争取更及时的治疗。

3. 农作物分析

在农业领域，人工智能有土壤检测、病虫害防护、产量预测等应用，例如汇总并利用各类数据，包括高分辨率的空中卫星和无人机图像、设备数据及天气等。在作物生长期间，智能分析平台会向农民的手机发送警报通知，如需要重新种植问题农作物、辨别抗病杂草和缺乏营养的农作物，以及可能影响收获的干枯率。

4. 智能客服

许多企业和机构都采用了智能客服，如电商平台、银行等。智能客服能够快速响应用户的咨询，自动解答常见问题。以电商平台为例，当用户询问商品的发货时间、退换货政策等问题时，智能客服可以瞬间给出准确答案，减轻了人工客服的工作压力，提高了客户服务的效率和响应速度 。

拓展知识

影视教学片段：AI 如何改变未来生活

你是否想过，人工智能将如何重塑我们的衣食住行？从清晨唤醒你的智能家居到危急时刻拯救生命的纳米医疗，AI正在悄然改写人类文明的运行逻辑。本教学片段通过沉浸式场景，展现AI在医疗、艺术、交通等领域的颠覆性应用，揭示人与技术共生共演的未来图景。请扫二维码，观看未来生活实景模拟。

思考：当机器能理解蒙娜丽莎的微笑弧度时，人类将如何重新定义创造力与情感？

作业与测评

1. 采用一种你常用的大模型来获取与食堂推荐系统相关的文献、文档、概念设计等资料，并分析其使用的数据、算力、算法和场景。

2. 在食堂智能推荐系统中，除了个性化推荐，还可以引入哪些AI技术来提高食堂的运营效率？

项目 2
人工智能伦理与责任

案例导学 校园监控系统的讨论

场景描述

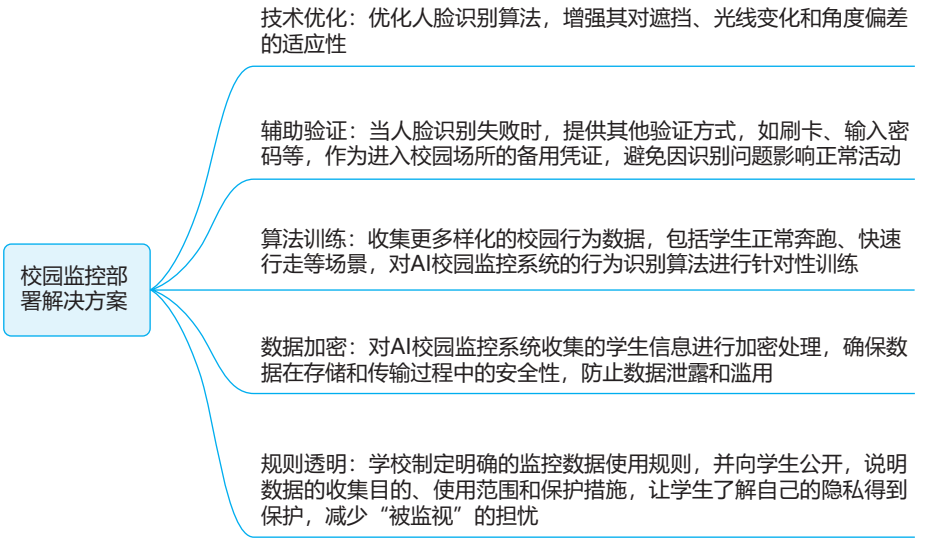
高职院校打算部署AI校园监控系统，但面临以下问题

- 人脸识别的局限性：学生若戴口罩、帽子，或光线过暗、角度偏差，可能导致人脸识别失败，引发争议。
- 正常行为误判：比如学生在走廊快速奔跑（可能只是赶时间）被误判为“追逐打闹”等，导致误判
- 隐私与伦理争议：可能让学生产生“被监视”情绪，引发对“心理隐私被侵犯”的担忧，甚至产生抵触心理

解决思维

问题：隐私侵犯、误判问题、心理压力 → 数据：面部信息、行为信息、考勤信息、情绪数据等 → 算法：人脸识别算法、行为分析算法 → 工具：摄像头、数据存储服务器、AI监控系统平台 → 应用：校园安全、教学质量、学生管理、考勤管理

解决方案



高职院校 AI 校园监控系统的解决方案需要基于伦理原则构建技术与管理体系。在数据安全与隐私保护方面，应明确限定公共区域数据采集范围，对人脸识别等敏感信息进行加密脱敏存储，同时建立用户知情授权机制，允许师生自主选择是否参与数据采集；在算法优化层面，通过数据增强平衡训练样本，引入人工审核环节降低误判率，并开通便捷申诉渠道处理识别争议，从技术源头减少伦理风险。

此外，还需要从管理与监督维度落实全方位治理。合理规划监控时段与范围，设置午休、课间等休眠期以缓解学生心理压力，配套开展心理讲座与沟通机制；成立由技术、伦理、法律专家及师生代表共同组成的伦理审查委员会，引入第三方机构对数据使用全流程进行独立监督，通过透明化的审查流程与动态调整机制，确保监控系统在提升校园管理效率的同时，切实保障师生权益与伦理合规。

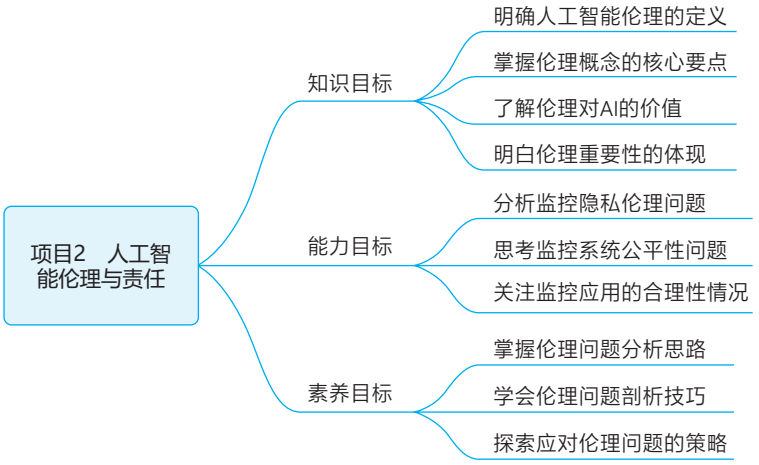
学生讨论

1. 人工智能在各个领域的应用越来越广泛(如学生用 AI 生成课程作业、教师用 AI 辅助备课)，可能会引发哪些关于内容版权与知识产权的伦理问题？请结合本项目中的相关案例和原则，提出至少两种应对策略和建议。

2. AI系统可能继承人类创造者的偏见和训练数据中的偏见，导致歧视性、偏见或不公平的输出。例如在招聘、司法等领域，AI的决策可能会在无意中歧视某些群体，损害社会公正。对此该如何应对？

3. 当AI做出错误决策导致他人受到伤害时，谁应该承担责任？例如自动驾驶汽车造成事故，汽车制造商、软件开发商、“司机”等谁应该承担责任？

学习导图



知识学习

人工智能伦理是指在人工智能技术的研发、应用和管理过程中遵循的道德原则及行为规范。它旨在确保人工智能技术的发展和应用符合人类的价值观及社会利益，避免对人类社会、个人权益和自然环境造成负面影响。

2.1　人工智能伦理的重要性

保护个人权益：随着人工智能技术的广泛应用，大量个人数据被收集和处理，如人脸识别、情绪检测等。人工智能伦理强调保护个人隐私和数据安全，防止数据泄露和滥用。

维护社会公平：人工智能系统的决策可能会对不同群体产生不同的影响。人工智能伦理要求确保算法的公平性，避免对特定群体的歧视，维护社会的公平和正义。

促进技术可持续发展：人工智能技术的快速发展会对社会和环境产生深远影响。人工智能伦理规范并引导技术的发展方向，确保其符合可持续发展的要求，避免对社会和环境造成不可逆转的伤害。

2.2　人工智能伦理的核心原则

2.2.1　尊重隐私原则

数据收集边界：明确数据采集的目的与范围，仅收集与任务直接相关的信息(如校园监控

仅采集行为分析必需的影像，而非全场景录音)。

数据保护措施：采用加密存储、匿名化处理等技术(如人脸识别数据须脱敏存储)，禁止未经授权的共享与滥用。

用户知情权：向用户明确告知数据使用方式，提供“拒绝采集”的选择权(如学生可申请不参与情绪检测系统)。

经典案例：校园监控禁用宿舍区域摄像头。要采用数据最小化收集的原则和加密脱敏技术来处理数据。

2.2.2　公平公正原则

算法无歧视性：避免因数据偏差导致对特定群体的偏见(如招聘AI须平衡性别、种族等样本比例，参考亚马逊招聘系统优化案例)。

决策透明性：向用户解释算法逻辑(如推荐系统须说明推荐依据)，避免“黑箱”决策导致的不公平。

动态纠偏机制：定期审查算法输出，根据反馈调整模型(例如，若校园监控系统误判率超阈值，则须重新训练数据)。

经典案例：亚马逊招聘AI系统的性别歧视。历史数据中男性程序员占比过高，导致亚马逊开发的招聘AI系统对女性候选人产生歧视。最终，通过数据平衡技术将女性候选人通过率提升了18%。

2.2.3　责任明晰原则

多方责任划分：明确开发者(确保算法安全)、使用者(合规应用技术)、监管者(制定伦理规范)的权责边界。

可追溯机制：建立技术日志，记录数据处理与决策过程，便于问题溯源(如自动驾驶事故须回溯算法参数与传感器数据)。

经典案例：波士顿儿童医院AI误诊事件。2024年，波士顿儿童医院的AI系统将罕见病诊断为常规肺炎，导致患儿错过最佳治疗期。法院判决算法开发者、医院和医生分别承担45%、30%和25%的责任。

2.2.4　透明可解释原则

技术流程公开：向用户说明 AI 系统的功能边界与局限性(如校园监控须公示“行为识别准确率为 90%，可能存在误判”)。

决策逻辑可视化：通过简易图表或文字解释输出结果(如大模型生成新闻稿时，标注数据来源、生成逻辑等)。

2.2.5　可持续发展原则

资源与环境考量：优化算法效率，减少算力消耗(如用轻量化模型替代重型架构，降低碳排放)。

社会影响评估：在技术落地前评估对就业、伦理的长期影响(如“萝卜快跑”配套了失业人员再培训计划)。

经典案例：百度“萝卜快跑”无人驾驶车引发就业替代焦虑。百度的无人驾驶出租车服务“萝卜快跑”引发了公众对就业替代的担忧，同时引发了社会对AI技术侵权的争议。

2.2.6　人类主导原则

人机权责优先级：AI 决策需要服从人类监督(如医疗 AI 的诊断结果必须经医生审核)。

伦理底线约束：禁止 AI 技术突破法律与道德红线(如严禁用 AIGC 生成侵权内容或虚假信息)。

经典案例：AI换脸技术滥用，被用于制作虚假视频。

通过以上原则的落地，可在技术效率与伦理风险间建立平衡，推动 AI 技术的正向发展。

2.3　人工智能伦理的核心问题与场景

人工智能伦理的核心问题与场景涵盖身份真实性、内容版权与知识产权、教育与学术研究、隐私权与系统偏见等维度。这些挑战凸显技术发展须与伦理治理同步：既需要通过立法明确界限(如严禁非同意换脸)，也需要推动“技术向善”的行业自律(如伦理审查委员会)。

2.3.1　身份真实性 (涉及人格权、社会信任危机)

1. 数字分身技术

伦理困境：未经授权创建逝者或生者的数字分身，侵犯肖像权、人格尊严；引发“自我同一性”哲学争议(如数字分身是否代表真实个体)。

案例：使用AI技术“复活”已故明星用于商业广告，家属起诉侵犯死者人格权；学者数字分身授课引发“知识传承主体归属”争议。

2. 深度伪造滥用

伪造视频或换脸诈骗：利用AI换脸和拟声技术冒充亲友诈骗(如诱导老人转账)。

不雅内容生成：“AI脱衣”技术批量生成虚假色情图像，侵害普通人的名誉权。

社会影响：破坏“眼见为实”的社会信任基础，2023年全球深度伪造诈骗案增长30%。

2.3.2　内容版权与知识产权

1. 作品版权归属争议

AI生成内容版权：AI绘画、写作工具(如Midjourney、ChatGPT)产出的作品是否享有著作权？法律对此尚无定论，美国版权局曾驳回AI生成漫画的版权申请。

训练数据侵权：大模型使用受版权保护的文本或图像进行训练，被指控“系统性抄

袭”(如艺术家抵制AI头像生成器)。

2. 虚假新闻与信息操纵

伪造新闻生产：AI批量生成虚假舆情(如伪造“×××带货直播”)，加剧社会分裂。

算法推荐偏见：社交媒体算法推送同质化信息，形成“信息茧房”，强化群体偏见(如政治极端化)。

2.3.3　教育与学术研究

1. 教学颠覆与责任模糊

AI替代教师角色：数字分身教师提供24小时辅导，削弱真实师生的情感联结，淡化教育的人文关怀。

学术诚信危机：学生用ChatGPT代写论文，检测工具(如GPTZero)识别率仅92%，学术评价体系受到冲击。

2. 科研伦理失范

数据造假与署名争议：篡改实验数据训练AI模型；AI辅助研究引发成果归属纠纷(如“双签字制度”能否保障责任明晰)。

算法偏见渗透研究：医疗AI诊断系统对黑人患者的漏诊率高达34%，反映训练数据的社会偏见。

2.3.4　隐私权与系统偏见

1. 数据隐私侵犯

AI伴侣类应用：未经授权调用摄像头、通信录，精准追踪用户行为(如AI伴侣说出用户未公开的聊天内容)。

生物信息泄露：人脸识别数据库遭黑客拖库，导致数万人的生物信息流入黑市。

2. 算法歧视与公平性缺失

系统性偏见：招聘AI筛选简历时降低女性评分；信贷模型对部分群体的授信率低30%。

教育资源配置不公：智能教育系统依赖硬件设备，加剧城乡教育资源差距。

2.4　解决人工智能伦理问题的策略

针对人工智能应用中的伦理挑战，需要从技术、制度、教育等多维度构建系统性解决方案，以下结合案例与实践经验展开说明。

2.4.1　技术层面：从源头降低伦理风险

1. 算法优化与鲁棒性设计

数据均衡技术：通过过采样、欠采样等方法修正训练数据偏差(例如，亚马逊招聘 AI 通过平衡性别样本降低歧视率)。

可解释性算法：采用因果推断模型或可视化工具(如 LIME)，让 AI 决策逻辑可追溯(例如，医疗AI须标注诊断依据的影像特征)。

动态纠错机制：设置误判阈值触发模型重训练(例如，校园监控系统若连续 3 次误判正常行为，则自动启动数据迭代)。

2. 隐私保护技术应用

差分隐私算法：收集数据时添加随机噪声，确保个体信息不可追踪(例如，统计食堂消费数据时模糊具体用户ID)。

联邦学习框架：数据在本地即可完成模型训练(例如，多所医院联合训练医疗模型时，无须共享原始病历)。

2.4.2　制度层面：构建全流程监管体系

1. 伦理审查机制

事前评估：技术落地前须通过伦理委员会审查(例如，校园监控系统须提交《隐私影响评估报告》，参考欧盟发布的《可信 AI 评估清单》)。

事中监控：建立实时风险预警系统(例如， AI 客服若识别到敏感问题，自动切换至人工审核)。

事后追责：明确事故责任划分标准(例如，波士顿儿童医院 AI 误诊事件中，按开发者45%、医院30%和医生25%的比例追责)。

2. 法律法规与行业规范

国家层面：遵循我国发布的《新一代人工智能伦理规范》，禁止数据滥用(例如，校园监控不得将情绪数据用于非教学目的)。

行业标准：参考 IEEE发布的《人工智能伦理设计准则》，制定技术实施细则(例如，自动驾驶须公开碰撞测试数据)。

2.4.3　管理层面：透明化与权责划分

1. 用户知情与参与

透明化告知：用通俗的语言说明 AI 功能边界(例如，智能推荐系统须提示“基于历史消费数据生成建议”)。

用户控制权：提供数据删除、功能关闭选项(例如，学生可选择退出校园情绪监测系统)。

2. 多方协作治理

跨领域协作：组建由技术专家、伦理学家、法律人士构成的治理小组(例如，学校部署AI 系统时，须吸纳学生代表参与方案设计)。

企业责任落实：科技公司须设立伦理合规部门(例如，百度 “萝卜快跑” 团队须定期提交就业影响评估报告)。

2.4.4　教育层面：提升伦理意识与能力

1. 技术从业者培训

伦理课程必修：AI开发者须通过伦理考试(例如，掌握数据偏见检测方法)，参考 Kaggle 伦理认证体系。

案例警示教育：以亚马逊性别歧视、波士顿误诊等案例为教材，强化责任意识。

2. 公众科普与素养教育

校园伦理课程：在高职 AI 通识课中设置伦理模块(例如，要求学生设计 “负责任的课堂专注度检测系统”)。

社会宣传活动：通过短视频、讲座普及AI 伦理知识(例如，解释AI 换脸技术的侵权风险)。

2.4.5　典型场景策略应用：校园 AI 监控系统

1. 隐私保护策略

仅采集公共区域影像，禁用宿舍、卫生间等敏感区域监控。

人脸识别数据加密存储，且仅用于考勤统计，不与其他系统关联。

2. 误判应对策略

建立学生申诉渠道：若对考勤记录有异议，可提交人工复核(例如，通过小程序上传打卡凭证)。

定期优化算法：每学期采用不同性别、体态的学生数据测试行为识别模型，确保误判率＜3%。

3. 心理干预策略

提前公示监控范围与目的，降低学生心理压力。

设立“监控休眠时段”(如下午课后 30 分钟)，给予学生放松空间。

2.4.6　应急响应与风险预案

伦理风险清单：梳理技术应用中的潜在问题(如 AIGC 生成内容的版权纠纷)，制订对应解决方案。

紧急熔断机制：当AI系统出现重大伦理事故(如数据大规模泄露)，立即暂停服务并启动调查。

通过以上策略的协同实施，可在技术创新与伦理规范间找到平衡点，推动 AI 技术安全、可持续地服务于社会。

2.5　伦理问题分析

在人工智能应用中，伦理问题的识别与解决需要依托系统化的分析框架和实用工具，以下结合案例与行业实践，从方法论与工具库两方面展开说明。

2.5.1　伦理问题分析方法论

1. 多维度评估框架

PESTLE 模型延伸：从政治(policy)、经济(economic)、社会(social)、技术(technical)、法律(legal)、伦理(ethical)维度分析影响。例如，校园监控系统的伦理评估需要考虑社会对隐私的敏感度(社会维度)、数据保护法规(法律维度)及学生心理影响(伦理维度)。

利害相关者分析法：识别技术影响的所有群体(开发者、用户、监管者等)，评估各方权益冲突。例如，AI 客服系统须平衡企业效率(减少人工成本)与用户知情权(明确“你正在与 AI 对话”)。

2. 伦理风险矩阵

风险等级划分：基于发生概率和影响程度将伦理问题的应对优先级分为高、中、低(见图2-1)。

表 2–1　风险等级划分

风险类型	案例 (校园监控)	发生概率	影响程度	应对优先级
隐私侵犯	人脸识别数据泄露	中	高	高
误判导致不公	正常行为被误判为危险行为	低	中	中
心理压力	学生因监控产生焦虑	高	低	中

3. 伦理决策流程

问题识别：通过场景模拟发现潜在风险，例如假设AI换脸技术被用于伪造校园通知。

原则对照：匹配伦理基本原则，例如判断该行为是否违反尊重隐私原则。

方案权衡：比较不同解决方案的利弊，例如数据匿名化与完全禁止采集。

动态调整：根据技术迭代与反馈持续优化，例如每学期更新校园监控的伦理评估报告。

2.5.2　伦理分析实用工具

1. 政策与标准工具

《新一代人工智能伦理规范》自查清单：检查点为数据采集是否获得用户明确同意，例如校园监控须向学生发放《知情同意书》。

欧盟发布的《可信 AI 评估清单》：核心指标有算法可解释性、隐私保护措施、社会影响评估。

2. 技术检测工具

(1) 数据偏见检测工具。

AIF360(IBM)：分析训练数据是否存在性别、种族等偏差，例如检测招聘 AI 的历史数据中女性样本占比。

Fairlearn(微软)：评估算法决策的公平性，提供纠偏建议，例如调整参数使不同群体的通过率差异＜5%。

(2) 隐私保护工具。

Opacus(PyTorch 差分隐私库)：在模型训练中添加噪声，防止数据溯源，例如统计食堂消费数据时注意保护个体隐私。

TensorFlow Federated：支持联邦学习，数据在本地即可训练模型，例如多校区联合优化考勤算法时，无须共享原始人脸数据。

3. 可视化分析工具

伦理决策思维导图工具：使用 XMind 或 ProcessOn 绘制伦理分析流程，关联案例与解决方案，例如将亚马逊性别歧视案例与数据均衡技术方案关联。

风险热力图工具：用 Excel 或 Python 绘制风险矩阵图，直观展示伦理问题的优先级，例如将 “隐私侵犯” 标注为红色高风险区域。

4. 合规性评估工具

伦理审查流程管理系统：参考学校部署 AI 系统时的审批流程，设置“数据安全评估”“学生代表意见征集”等必填环节。

2.5.3　典型场景分析示例：校园 AI 监控系统

1. 伦理问题识别工具应用

利害相关者分析法：引出利害相关者，如学生(隐私、心理压力)、教师(教学管理效率)、学校(安全责任)、技术供应商(商业利益)。

用风险矩阵定位高风险点：人脸识别数据若未加密存储，发生泄露的概率为中等，影响程度为高，须优先解决。

2. 工具落地策略

技术工具：使用 AIF360 检测行为识别算法是否对特定体型学生存在误判偏见。

管理工具：参照《新一代人工智能伦理规范》制定《数据泄露应急预案》，明确追责流程。

2.5.4　工具整合与实践建议

1. 工具链搭建

初级应用：用 Excel 制作伦理风险清单和用XMind 梳理分析流程。

进阶应用：利用零代码平台(如明道云、简道云)的表单引擎与工作流引擎构建可视化伦

理数据治理中台。在数据采集端，通过拖曳组件生成《人脸数据采集申请表》，设置 “采集区域” “数据用途” 等必填字段，并关联电子签章功能实现师生知情授权；在数据处理层，配置自动化规则，当人脸识别数据录入系统时，自动触发脱敏处理流程。

2. 工具使用培训

对高职学生的教学建议：用零代码平台分析课堂专注度检测算法的公平性，提交检测报告。

通过方法论与工具的结合，可将抽象的伦理问题转化为可量化、可操作的解决方案，提升 AI 系统的伦理合规性。

实训任务

任务 1　AI 推荐系统公平性优化

【任务场景】

校园食堂 AI 点餐系统因历史数据中男生点餐记录占比 60%，导致女生偏好的低糖菜品推荐频率低 18%。

【任务要求】

(1) 分析该场景存在的伦理问题(须关联公平公正原则)。

(2) 零代码平台操作示例(以适己科技零代码平台为例)。使用数据表单创建 “食堂消费记录” 表单，通过 “关联字段” 功能，将消费记录与学生档案表(含性别信息)自动匹配，补全数据。

(3) 工作流配置。设计 “数据平衡流程” ：当录入新消费数据时，系统自动检测性别分布，若某一性别样本占比超过 55%，触发过采样或欠采样规则(如复制女生低糖菜品记录)。配置 “推荐算法执行流” ：根据性别、历史消费偏好、菜品糖分标签，调用预设的权重公式生成推荐列表，并标注推荐理由。

界面效果见图2-1和图2-2。

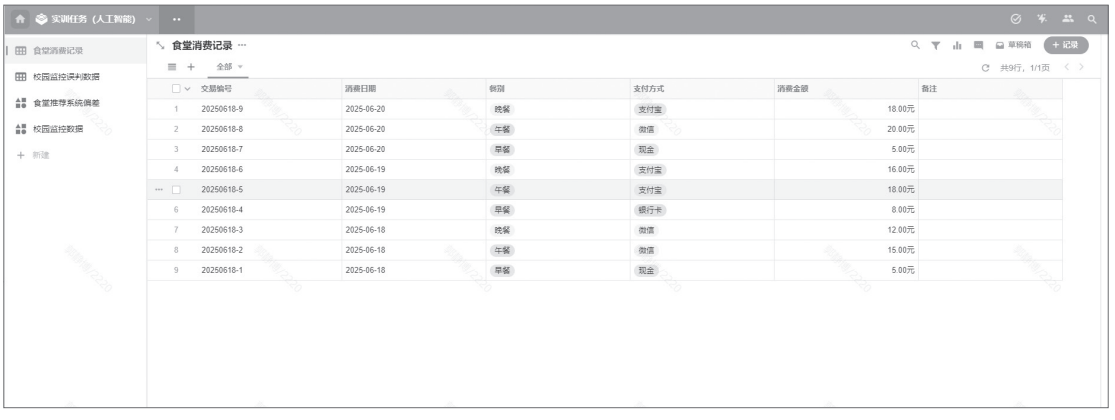


图 2-1　食堂消费记录界面

图 2-2　食堂推荐系统偏差界面

任务 2　检测校园监控系统数据偏见

【任务场景】

校园监控系统的行为识别算法可能因训练数据分布不均产生偏见。例如，监控系统误判某体型学生“打架”的概率偏高。

【任务要求】

(1) 数据整理与分类，导入原始数据——校园监控误判数据(见图2-3)。

(2) 用Excel公式计算不同群体的误判率。

(3) 用图表将偏见可视化(见图2-4)。

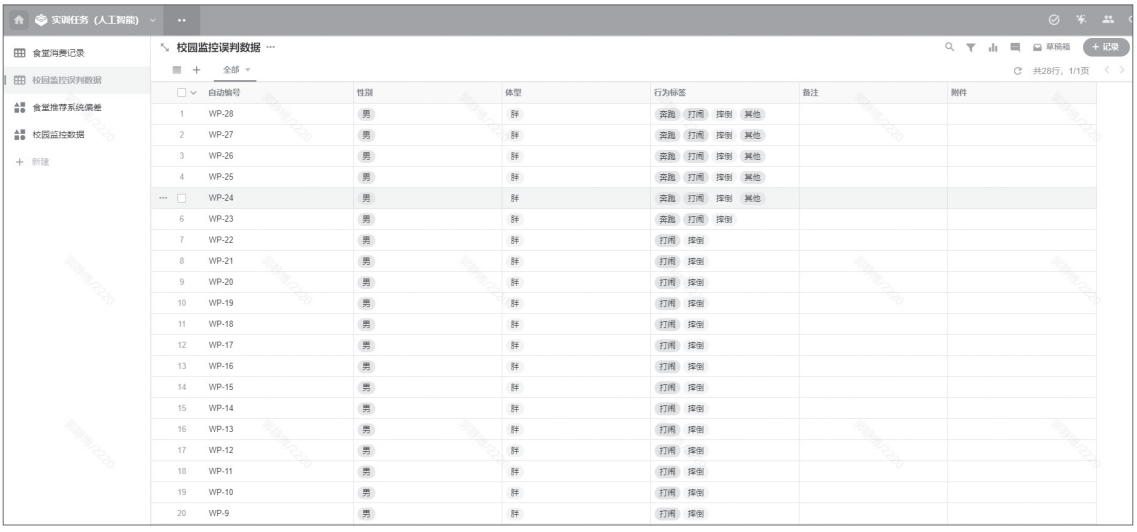


图 2-3　校园监控误判数据

图 2-4　图表数据可视化

拓展知识

国内外涉及人工智能伦理的主要法律法规参见表2-2和表2-3。

表 2–2　我国涉及人工智能伦理的主要法律法规

法规名称	生效时间	核心要点	典型应用场景
《生成式人工智能服务管理暂行办法》	2023年8月	① 训练数据合法合规 ② 生成内容标识，安全评估 ③ 禁止歧视性算法	校园AIGC应用(如AI写作助手)
《互联网信息服务算法推荐管理规定》	2022年3月	① 禁止“大数据杀熟” ② 保障用户选择权 ③ 定期审核算法模型	智慧校园推荐系统(如课程推送)
《个人信息保护法》	2021年11月	① 数据采集须“单独同意” ② 敏感信息(人脸等)特殊保护 ③ 跨境传输安全评估	教室人脸考勤系统
《数据安全法》	2021年9月	① 数据分级分类保护 ② 重要数据备份加密 ③ 操作日志留存6个月以上	校园监控数据存储
《新一代人工智能伦理规范》	2021年9月	① 人类主导原则 ② 公平、普惠 ③ 隐私保护	AI教学设计准则

表 2–3　国际上涉及人工智能伦理的主要法律法规

国家、地区或组织	法规/框架	特色条款	对我国的启示
欧盟	《人工智能法案》(2024年通过)	① 风险四级分类： ● 禁止类 ● 高风险类 ● 有限风险类 ● 最小风险类 ② 人脸识别公共场所有限使用	中国高职智慧校园系统须参考“高风险类”管理
美国	《人工智能权利法案蓝图》(2022年10月)	① 算法歧视检测 ② 人类替代选项(可拒绝AI服务) ③ 数据隐私保护	建议校园系统设置“关闭AI监控”按钮
经合组织	《AI原则建议》(2019年修订)	① 包容性增长原则 ② 透明可追溯要求	我国《新一代人工智能伦理规范》的主要参考来源

作业与测评

1. 列举校园AI监控可能涉及的两个伦理问题(如隐私泄露、误判风险)，并各写一条解决办法。

2. 说明AI推荐系统收集学生数据时应遵守的两条伦理原则(如尊重隐私、公平公正)，举例解释。