

青少年人工智能伦理教育的探索及启示

——以 MIT 为例

李 艳,朱雨萌,樊小雨

(浙江大学,浙江 杭州 310058)

【摘要】随着人工智能行业应用的不断深入,其伦理问题日益凸显。《义务教育信息科技课程标准(2022 年版)》明确指出要面向中小学生开展有关人工智能伦理的教育。鉴于国内外面向青少年的人工智能伦理教育尚处于起步阶段,介绍美国麻省理工学院(MIT)自 2018 年起在该领域的探索能为国内中小学校人工智能伦理教育教学实践提供一些启示。MIT 青少年人工智能伦理教育涵盖小学、初中和高中三个学段,以初中和高中为主,涉及家庭、学校和社会三个场域,有专题式和内嵌式两种类型,形式上包括与中小学合作的线下 STEAM 选修课程、线上/线下的夏令营活动以及面向社会招募学生的混合式工作坊。梳理和反思 MIT 青少年人工智能伦理教育实践可以得到以下三方面启示:(1)高校可以与中小学合作开展青少年人工智能伦理教育的内容设计与教师培训;(2)中小学多学科教师可以尝试开展内嵌式青少年人工智能伦理教育设计与实践;(3)“家校社”应携手共建多样态多场域本土化青少年人工智能伦理教育生态。

【关键词】人工智能教育;人工智能伦理教育;青少年;MIT

【中图分类号】G43

【文献标识码】A

【文章编号】1001-8700(2024)01-0003-11

DOI:10.13927/j.cnki.yuan.20240022.001

当今社会正处于以“人-人-机”三角关系主导的弱人工智能机器时代^[1],人工智能具有算法不透明、难解释、自适应等特征^[2],也正因如此,人工智能的行业应用也存在算法偏见、隐私暴露、各种歧视等伦理议题^[3-4]。智能技术在为个人和行业带来诸多便利的同时,也在为个人、机构以及国家带来前所未有的伦理风险与挑战^[2]。为应对人工智能技术带来的伦理挑战,2021 年 9 月 25 日,国家新一代人工智能治理专业委员会发布了《新一代人工智能伦理规范》,将提升伦理素养作为六项人工智能基本伦理要求之一,明确指出人工智能使用者应积极学习和普及人工智能伦理知识,主动开展或参与人工智能伦理问题讨论^[5]。2022 年 11 月 16 日,我国向联合国递交了《中国关于加强人工智能伦理治理的立场文件》,强调“人工智能治理应坚持伦理先行”,指出“各国政府应重视公众人工智能伦理教育”^[6]。科普人工智能伦理知识成为人工智能技术变革社会下的当务之急。

2022 年,教育部新修订的《义务教育信息科技课

程标准(2022 年版)》(简称“新课标”)明确指出,信息科技课程旨在“培养科学精神和科技伦理,提升自主可控意识,培育社会主义核心价值观,树立总体国家安全观,提升数字素养与技能”^[7]。中小学生在义务教育阶段如能形成科学的、正确的人工智能伦理价值观,能认知智能技术非理性应用可能带来的负面效应,则对整个智能社会的健康发展具有重要意义^[8]。然而,我国当下面向青少年的人工智能伦理教育存在教学资源稀缺、可借鉴课程案例不足等问题^[9]。现有科技伦理教育和人工智能伦理教育的开展主要集中在高校^[10-12],其往往要求学生具备较强的思辨能力,能够理解较为抽象的概念,难以直接下沉到面向青少年的基础教育学段^[13]。

近年来,学界开始兴起对青少年人工智能伦理教育的探讨。例如,Lin 等使用增强现实技术,面向我国六年级学生开展了为期两周的人工智能伦理知识教学与伦理困境讨论活动,发现使用增强现实技术能够更好地提升学生人工智能伦理意识和高阶思维倾

【基金项目】科技创新 2030 重大项目子课题“人机协同的学习社群建构与支持技术”(编号:2022ZD0115904)。

【作者简介】李艳,教授,浙江大学教育学院博士生导师;朱雨萌,浙江大学教育学院博士研究生;樊小雨,浙江大学教育学院硕士研究生。

向^[14]。然而,我国面向青少年人工智能伦理教育仍处于起步阶段,需要长周期打磨课程设计、优化课程资源和沉淀教学经验。美国麻省理工学院(Massachusetts Institute of Technology,简称 MIT)是全球最早关注青少年人工智能伦理教育的高校之一。自2018年以来,MIT不断探索青少年人工智能伦理教育的课程设计与实践,其开发的青少年人工智能伦理课程先后被德国、西班牙、韩国等17个国家使用^[15]。本文尝试系统梳理MIT面向青少年开展的多项人工智能伦理教育实践项目,分析不同项目的内容、形式、贡献、编程要求与课程时长等,介绍项目团队、设计、实施结果与发现,以期为更好地设计与开展我国青少年人工智能伦理教育带来启示。

一、MIT 面向青少年的人工智能伦理教育实践概述

MIT 媒体实验室(MIT Media Lab)是MIT面向青少年人工智能伦理教育的积极推动者。其下属个人机器人小组(Personal Robots Group,简称 PRG)是辛西娅·布雷泽尔(Cynthia Breazeal)教授于2003年创立的团队,该小组于2018年开始探索青少年人工智能伦理教育。近年来,该团队运用教育学、心理学、设计学、美学、复杂计算等多学科知识与方法,以工作坊、选修课等方式面向青少年群体开发了系列人工智能教育项目,旨在培养青少年人工智能素养,引导青少年初步了解并思考人工智能及其利弊,掌握人机交互

技能,积极参与人工智能相关活动,形成正确的科技伦理观,为未来设计和开发合乎伦理的人工智能技术做好准备。

2021年,MIT媒体实验室、MIT苏世民计算学院以及MIT开放学习中心联合启动了“对社会赋权和教育负责任的人工智能”项目(Responsible AI for Social Empowerment and Education,简称 RAISE)。RAISE以PRG自2018年来的工作为基础,以PRG为主导,汇集了计算机科学与工程、媒体比较与写作、机械工程、管理学、脑科学和认知科学以及教育学等多学科的专家和学生,共同探索人工智能时代的创新教学实践。针对人工智能技术潜在的滥用和算法歧视问题,RAISE强调人工智能使用中的公平性问题,期待通过人工智能及其伦理教育让K-12学生、社会劳动者和终身学习者负责地使用人工智能技术,在智能社会中拥有平等权利。

PRG和RAISE主导的青少年人工智能及其伦理教育实践涵盖小学、初中和高中三个学段,包括家庭、学校和社会三个教育场景,形式包括与中小学合作的STEAM选修课程、线上/线下夏令营以及面向社会招募学生的工作坊。根据人工智能伦理教育内容在项目中的形式,MIT面向青少年的人工智能伦理教育项目可分为内嵌式和专题式两种类型。整体而言,MIT面向青少年人工智能伦理教育对象以初中生为主,课程对学生编程要求门槛低,内容难度根据学段逐渐上升(如图1所示)。

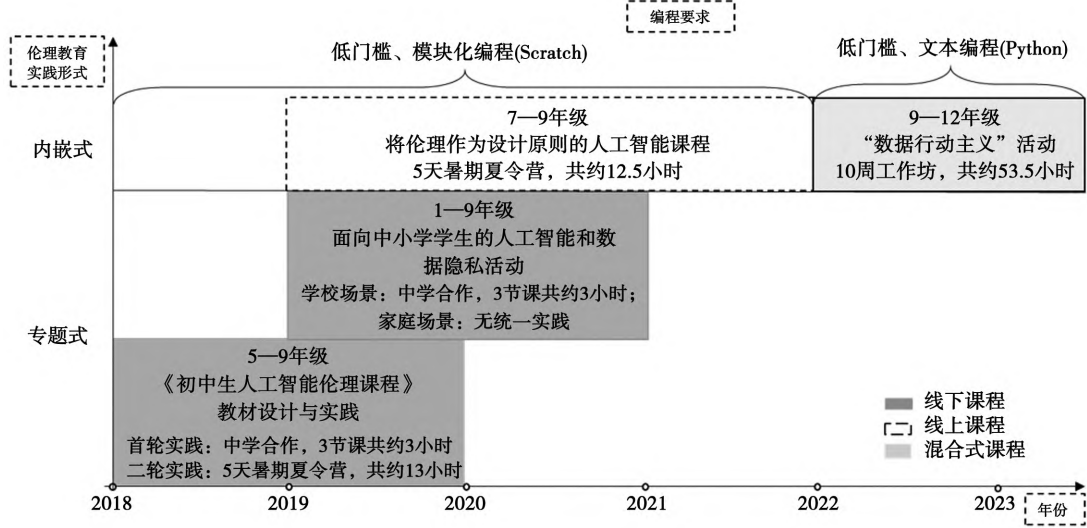


图1 2018—2023年MIT面向青少年的人工智能伦理教育项目概览

二、MIT 面向青少年的专题式人工智能伦理教育实践

(一)“初中生人工智能伦理课程”的设计与实践
PRG 硕士研究生布莱克利·佩恩(Blakeley H.

Payne)和博士研究生丹妮拉·迪保拉(Daniella Dipola)主导设计了“初中生人工智能伦理课程”,课程向青少年科普人工智能基本原理,让青少年能够了解所有人工智能系统都是为利益相关者服务而非完全客观中立的特点,分析人工智能系统中反映利益相关者

价值的设计特征以及这样的设计对利益相关者/其他消费者的影响。该课程的教材经过两轮课程实践并于2019年正式发布,已被译成德语、韩语和葡萄牙语等多国语言版本^[16]。

1. 课程设计与首轮实践

课程设计团队通过采访美国多所城乡小学、中学、高中的教师及教育管理者,明确了青少年人工智能及伦理教育是中小学教师共同关注但无法解决的“痛点”问题^[17],决定以不插电、低门槛及强调伦理的原则进行课程设计。课程的初次设计旨在通过三节课向中学生传授人工智能基本知识及其伦理风险,内容包括“监督式机器学习和算法偏见”“作为意见的算法”“伦理矩阵”“YouTube再设计”等^[18]。第一节主要介绍人工智能、算法偏差等基础概念。学生通过训练并对比“有偏见”和“无偏见”的猫狗分类模型,探索因数据集不同导致的算法偏见现象,并迁移到对人脸识别算法偏见的讨论。第二节课依托制作“最好”的三明治活动,向学生介绍“作为意见的算法”和“基于伦理矩阵的利益相关者价值分析”两个概念。学生通过讨论“最好”的不同标准(如最好吃的、最健康的等),意识到不同利益相关者(如吃三明治的儿童、父母或医生)对“最好”的价值取向不同。在第三

节课中,学生重新设计 YouTube(在线视频平台)的推荐算法原型。学生们利用伦理矩阵分析视频推荐算法利益相关者及他们对算法的价值倾向,并根据价值倾向确定推荐算法优先考虑的因素,根据伦理矩阵和推荐算法考虑因素给出推荐算法的原型。

2018年10月,PRG与宾夕法尼亚州的大卫·威廉姆斯中学(David E. Williams Middle School)合作,面向该校225名五至八年级没有编程基础和人工智能学习基础的青少年开展了“初中生人工智能伦理课程”的首轮实践。学生以选修课的形式参加了三次45分钟的授课活动。

2. 首轮实践效果与启示

首轮实践发现,学生们课后能够识别出身边更多样化的人工智能技术应用,也能意识到算法偏见等伦理问题的存在,并知晓人工智能技术存在潜在的不公平性。学生们在课程动手实践环节参与度最高。在第三节课中,学生将视频推荐算法利益相关者对算法的价值倾向归纳为盈利、教育、娱乐等不同类型(如图2左图所示)。学生根据伦理矩阵和推荐算法考虑因素给出推荐算法的原型(如图2右图所示),其特征包括“易见的家长控制功能”“供 YouTube 用户选择的多种推荐算法”等。

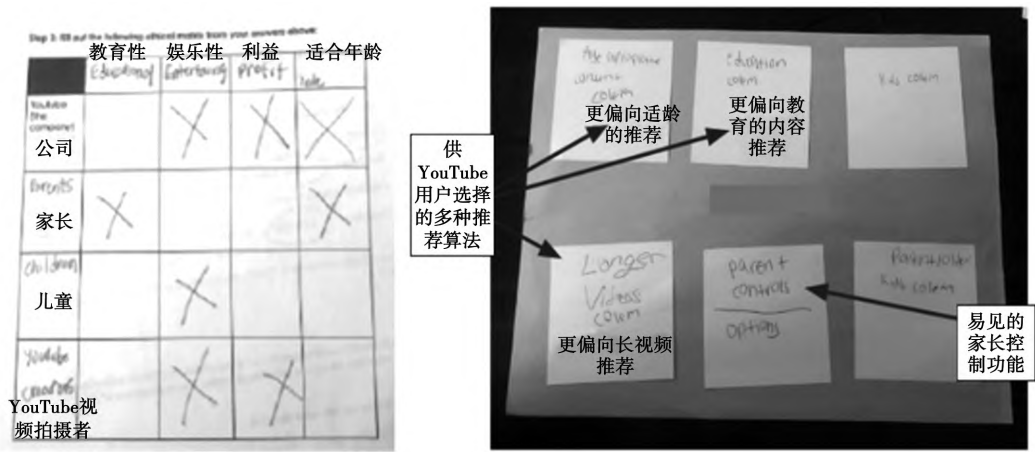


图2 首轮实践中学生对 YouTube 推荐系统的利益相关者分析(左图)及学生设计的“适合年龄的 YouTube”纸质原型(右图)

针对首轮实践效果,项目团队总结了该类活动的设计原则。首先,课程内容需关联初中生日常生活,以激发学生深入思考并讨论日常使用的人工智能技术及潜在的伦理问题。例如,对于面部识别的伦理问题,初中生们很难将机场监控技术关联到面部识别技术并展开讨论,但他们熟悉手机面部识别开锁(Face ID)场景,很容易展开深入讨论。再如,相较于自动驾驶技术,学生们更愿意重新设计他们熟悉的 YouTube,并识别利益相关者。相比于对伦理问题进行批判与讨论,青少年认为利用伦理矩阵进行设计能够收获更

多知识^[17]。其次,课程内容应包含对人工智能基础知识和概念的介绍,以帮助学生更好地识别人工智能技术并理解其工作原理。大多数学生将人工智能与机器人或语音代理混为一谈,且不能区分编程、算法、硬件和电子产品等概念。实践还发现,主流社交媒体影响了部分学生对“人工智能伦理”的理解。学生对于人工智能的担忧集中在强人工智能、隐私问题、对工作的影响以及人工智能错误运行的后果。因此,课程内容设计也要考虑社交媒体对学生人工智能及其伦理问题先验认知的影响。最后,实践发现课程需要

为小组合作预留更多时间并优化小组合作策略。学生们表示开展“重新设计 YouTube”的活动时间太短,而且,出现了个别四至五人小组由于很难在利益相关者群体的确认上达成共识而导致活动无法推进的问题,影响了活动的成效。后续,教师可以尝试预留足够活动时间,并针对小组活动出现的问题进行必要的干预或策略实施。

3. 教材发布与第二轮实践

项目团队于 2019 年完善和正式发布了该教材,修订后的教材提供了人工智能基础术语的定义及解释,将课程内容关联到学生熟悉的日常生活,并设计

了真实问题情景以激发学生的学习兴趣。修订后的教材共包含八项核心活动,包括 AI 宾果游戏、作为意见的算法、伦理矩阵、有监督机器学习和算法偏见、推想小说、YouTube 寻宝游戏、重新设计 YouTube 以及 YouTube 研讨会。八个活动内容呈现深入浅出、螺旋上升的特点(如表 1 所示)。

2019 年,教材以为期五天的夏令营形式在 MIT 校内开展了第二轮实践,面向 28 名五至九年级学生。学生们按照活动内容的安排参与了八项活动。两名研究生作为主讲教师,一名本科生和三名来自 STEM 项目教育公司的教师担任助教。

表 1 MIT“初中生人工智能伦理课程”活动概览

活动名称	时长	活动内容
AI 宾果游戏	30 分钟	学生两人一组填写关于人工智能系统的宾果卡,确定各个系统的预测内容以及所使用的数据集
作为意见的算法	45 分钟	在理解算法的含义后,学生需要撰写制作“最好”的花生酱和果冻三明治的算法,并探索“最好”的含义,反思这些观点是如何反映在算法中的
伦理矩阵	45 分钟	学生需确定关心花生酱和果冻三明治算法的利益相关者以及他们在算法中嵌入的价值理念。学生需要填写伦理矩阵,思考这些价值理念的重叠或冲突之处
有监督机器学习和算法偏见	3 小时	学生理解“分类”概念后,利用 Teachable Machine 了解监督式机器学习并建立猫狗分类器。在出现带有偏见的分类结果后,在教师提示下使用新数据集对分类器进行重新训练
推想小说	3 小时	学生与各种人工智能技术交互后进行创造性地写作,想象受到技术影响的群体以及该技术在未来会存在怎样的危害或利益
YouTube 寻宝游戏	30 分钟	学生需要识别 YouTube 中的各种人工智能算法(例如广告匹配算法、视频推荐算法、评论分类器等),确定各个算法预测内容及其所使用的数据集
重新设计 YouTube	4 小时	学生应用前几节课所学的知识,构建伦理矩阵以确定 YouTube 推荐算法目标;在纸上设计新版本的 YouTube 并展示新版本中满足利益相关者价值理念的功能
YouTube 研讨会	30 分钟	学生阅读《华尔街日报》中关于 YouTube 的摘要后,以学生为中心开启讨论,包括哪些利益相关者在 YouTube 应用程序拟议更改中最具影响力,以及诸如自动播放等技术设计是否应该存在等问题

第二轮实践效果整体上优于第一轮实践,课后调研显示,学生更加理解人工智能系统的基本机制,能够分析社会技术系统中的众多利益相关者,并能运用所学知识重新设计体现特定目的的视频平台(如图 3 所示)。在 YouTube 研讨会上,学生们表达了从视频平台被动用户转向自主消费者的意愿和倾向,大大提升了他们在人工智能技术应用中的自主性。

(二) 面向中小学生的“人工智能和数据隐私”系列活动设计

2019 年,PRG 的研究科学家斯蒂芬妮·阮(Stephanie Nguyen)和博士研究生丹妮拉·迪保拉(Daniella Dipaola)依托 RAISE 设计了面向中小学生的“人工智能和数据隐私”系列活动。该活动通过在家

庭和学校两个场景向青少年介绍数据和隐私议题,鼓励青少年对人工智能进行批判性思考,而非记背相关安全伦理知识^[19]。

1. 家庭场景:父母与青少年讨论人工智能伦理问题

项目团队认为,父母和青少年在家中讨论智能技术应用中具体争议的话题,一方面有助于亲子沟通,另一方面有利于培养青少年阅读能力、批判性思维和讨论当下技术议题的能力,从而在未来能批判性地分析技术设计的价值观、动机与特征。项目团队制作了包括准备、参与、反思与重复四个步骤的家庭讨论活动指南。该指南对这四个核心步骤进行了详细解释,同时还包含了活动设计的背景、目的、开展方式、讨论



图3 第二轮实践中学生为“娱乐和盈利”目的设计的 YouTube 纸质原型(左图)及学生为“合理利用时间”目的设计的 YouTube 纸质原型(右图)

主题示例(如表2所示)和讨论文章示例,选取文章包括“脸书(Facebook)禁止利用人工智能技术进行内容造假”“视频软件频道被指控有意模糊视频和广告界限”等新闻报道。

表2 “与孩子讨论技术和人工智能的四步骤”活动指南中的讨论主题示例

讨论主题	具体问题
提炼问题	文章中建议或概述的主要问题是什么? 这为什么是个问题?
解决方案	提出的主要解决方案是什么? 还有哪些替代方案?
影响	这项技术对谁有帮助? 这项技术可能会伤害谁?
利益相关者	哪些利益相关者(学生、教师、家长、青年组织者等)最可能参与到有关该技术的讨论中? 他们的观点可能是什么? 他们观点中可能的异同点是什么?
竞争对手	哪些技术与这项技术相似? 相似技术和这项技术有何不同?
完全禁止或合法	某一技术使用行为应该被完全禁止还是不应被禁止? 为什么?
权力与决策	谁应为技术潜在危害作出决策并负责? 政府或用户(普通人)是否享有该权力与决策?
重新设计技术	公司如何重新设计技术以减少危害? 哪个利益相关者应受益更多?

2. 学校场景:向中小学生学习人工智能与数据隐私

项目团队设计了三个在学校里向青少年(7—14岁)介绍数据隐私的活动(如表3所示)。活动一方面是让青少年学习与数据隐私相关的技术与基础概念,另一方面意在帮助监管机构明确面向青少年的技术产品和服务中需要考虑的隐私问题。对于前者,课程设计了向青少年介绍实用数据隐私及其保护的相关内容,激发青少年对数据隐私和技术领域的兴趣;对于后者,该课程通过调研青少年常用应用程序来了解他们的数据隐私文化和沟通偏好,探索其特有的群体特征对数据隐私文化的影响,为向公众科普数据隐私话题带来启示^[13]。

三、MIT 面向青少年的融入式人工智能伦理教育实践

(一)将伦理作为设计原则的人工智能课程设计与实践

前期实践证明,青少年只有在了解人工智能相关

基础知识(如数据、算法等)后,才能对人工智能伦理问题进行批判性思考与讨论。为此,PRG 尝试将伦理作为课程设计原则融入人工智能课程设计之中。2019 年至 2021 年,PRG 博士生萨菲纳·阿尔沙德·阿里(Safinah Arshad Ali)与丹妮拉·迪保拉(Daniella Dipaola)主导的项目团队根据主动学习、嵌入式伦理和低准入门槛这三个设计原则,先后设计、开发、实践并发布了三门面向中学生的人工智能课程,即“如何训练你的机器人”“创造性人工智能”“与人工智能共舞”(如表4所示)。三门课程的知识点与内容均参照《CSTA K-12 计算机科学标准》^①以及《K-12 人工智能教学指南》^②制定,课程希望达到三方面的学习目标:(1)学生能掌握基础的人工智能技术知识(定义、生活中人工智能系统的特征等),理解人工智能算法功能以及人类在创建人工智能系统中的作用。(2)学生批判性地思考人工智能系统的潜在利弊、相关伦理问题(如算法公平、隐私保护、以人为本的设计)及利益相关者对其的影响。(3)学生能将人工智能知识和

技能恰当地应用到实际场景中,考虑道德伦理内容, 能产品^[20]。
并发挥创造力、同理心和综合思考技能来设计人工智

表 3 在中小学开展的人工智能和数据隐私系列活动设计

活动名称	时长	概述	学习目标
神秘的视频 平台:数据 隐私课程	30—45 分钟	什么是数据和隐私? 在此活动中,学生观察来自神秘用户的 YouTube 主页,以主页推荐视频为线索画出角色画像,猜测角色爱好与住所。学生由此体验 YouTube 算法功能:根据用户观看的内容来猜测用户的信息	了解技术的作用;了解数据隐私;了解隐私的影响
设计“透明” 的广告:创 意工作坊	30—45 分钟	在此活动中,学生学习批判性地思考广告在生活中的意义、目的和影响。学生分小组制作日常用品广告,思考应该突出哪些元素或信息来让广告尽可能“透明”(不被观察到)	分辨恶意软件和广告软件;了解数字化平台中数据隐私的重要性;了解广告的意义、目的和影响;了解如何规避广告诈骗
在社交媒体 平台中设计 数据隐私 法案	60—90 分钟	此课程教授学生有关 YouTube 设计、数据隐私和公民基本道德规范。学生对在线许可及其挑战(如数据使用协议、用户隐私协议等)提出自己的设计和解决方案。本课程让学生反思和分享自己的隐私观点,并了解影响美国青少年使用媒体的方式的最重要的隐私法之一:儿童在线隐私保护规则	学习设计、数据隐私和公民知识。反思和分享自己与他人的隐私观点。了解数字时代最大的挑战是在线用户同意将数据披露给数字化平台

表 4 三门融入式人工智能伦理课程简介

		如何训练你的机器人	创造性的人工智能	与人工智能共舞
发布年度		2019	2020	2021
学习目标		理解有监督机器学习算法	理解生成算法及其应用	训练有监督机器学习模型
		辨别人工智能系统的利益相关者	预测算法的社会影响	了解人工智能系统的伦理和社会影响
		开发人工智能系统	通过人机协作创造作品	设计交互式 AI 系统
设计原则	主动学习	在“介绍图像分类”活动中,学生们利用不同数据集训练图像分类模型,以了解图像的不同特征对模型结果的影响	在“GAN 工作原理”活动中,学生通过游戏模拟 GAN 两个神经网络(生成器、判别器)之间的关系来理解 GAN 工作原理	学生通过“猜字谜”活动识别、描述和表现运动的显著特征,设计运动分类器以学习“特征选择”概念
	嵌入式伦理	所有学生完成并展示结课项目的伦理矩阵,分析关键利益相关者并分析项目的潜在利弊与影响	在“探索 GAN”活动中学生与基于 GAN 的各种技术工具互动,展示并讨论这些工具的潜在影响	学生通过“人工智能分类示例”活动了解图像分类的工作原理,思考算法偏见对世界的影响
	低准入门槛	没有预修条件,第一个编程活动使用了一系列脚手架让学生快速掌握机器人技术和编程概念,以在后续课程应用	考虑到 GAN 通常需要大量的算力,课程所用工具使用预先训练好的模型,学生仅需在网页端运行即可快速创建文本和图像	PoseBlocks 能够在网页端进行实时情感和图像识别。没有专用的 GPU、相机、麦克风的学生设备也可以使用 Pose-Blocks

项目小组通过邮件从美国各地(加利福尼亚、佛罗里达、佐治亚、新泽西、纽约、俄亥俄和德克萨斯)招募了 11 名教师。每门课程由不同教师负责。课程开始前,项目团队对教师进行在线培训,一方面让教师充分了解教材及其涉及的人工智能背景知识,另一方面鼓励教师结合自身经验及创意修改完善课程。由

于疫情原因,三门课程于 2020 年以在线夏令营的方式开展。项目小组通过在线会议软件 ZOOM 和在线学习平台谷歌课堂同步举办了四个为期五天的在线夏令营,每天开课 2—3 个小时,学生们通过在线协作工具(如谷歌表格等)、分组讨论室和在线聊天平台进行在线协作学习,教师实时查看学生作业,帮助其调

试代码并提供代码调试相关的图解指南和视频教程。

1. “如何训练你的机器人”课程

PRG 与 i2 Learning(一个提供 STEM 课程和教师专业发展培训的商业组织)合作开发了“如何训练你的机器人”课程。课程内容包括基于有监督机器学习的语音识别和图像识别算法、用于应用程序和机器人的人工智能模型、人工智能模型导致的伦理问题以及基于道德的人工智能设计原则。基于开源的 Scratch Blocks 代码库,项目团队开发了人工智能模块化编程平台。学生通过该平台可以将 Machine Learning for Kids(MLK)平台训练的图像识别模型、文本分类模型和控制机器人的编程模块集成在一起,利用基于道德的人工智能设计原则设计为人类服务的人工智能项目,并训练相应功能的机器人。

项目团队于疫情前开展了第一轮项目实践,对来自城市的一名数学教师以及来自农村地区的一名数学教师和一名科学教师进行为期两天的培训。三位教师在其各自有 Scratch 使用基础的班级中进行为期一周的课程实践。课后项目团队采访了三位教师并根据其反馈简化了机器人的设置操作,并用 Teachable Machine 平台代替了 MLK 平台以解除学生模型训练数量的限制^[20]。

第二轮实践由于疫情无法开展线下授课,项目小组将教师培训与项目进行了简化并以线上教学形式开展。项目团队将蓝牙机器人 micro:bit 邮寄给学生,29 名中小学生居家在线学习课程,并在教师的在线辅导下独立制作机器人。

实践结果发现,课程有效激发了学生们的创造力,学生创造的 24 个机器人作品中,有 14 个作品使用了机器学习算法,有 8 个作品使用了机器人编程技术但没有使用机器学习算法。学生作品设计应用场景包括娱乐(7 个)、助人(7 个)、医疗保健(5 个)、科学(3 个)和教育(2 个),其服务对象主要包括儿童及青少年(5 个)、家庭(2 个)和社区(2 个)。学生在问题选择、编程和构建机器学习模型等方面表现出色。绝大多数学生能够根据指定问题设计机器学习算法、寻找训练集和测试集数据并训练得到机器学习模型。学生们课后反映,课程围绕机器人内容开展大大增加了他们的学习兴趣,有助于理解人工智能原理^[21]。不过,学生们在设计机器人时容易忽略评价个人作品的机制和标准,并且缺少根据用户反馈迭代作品的途径。

2. “创造性人工智能”课程

该课程旨在向中学生介绍生成式人工智能技术的定义、基本原理以及如何人机协作创造数字媒体艺

术。基于项目团队开发的学习工具包,学生们可以基于生成对抗网络模型(Generative Adversarial Network,简称 GAN)生成多模态内容,思考机器创作与人类创作的异同及其潜在伦理问题。同时,学生们可以在课程中了解深度伪造(Deepfake)的概念,即由人工智能生成的、非常逼真的但现实生活中不存在的人或物体,并尝试鉴别人和深度伪造所生成内容,讨论虚假媒体信息带来的危害,并思考深度伪造生成信息的在线传播途径。

项目团队依据布鲁姆学习目标分类理论设计课程目标,通过四个主题和五个活动的设计将学习内容和学习目标进行细分:(1)活动 1:生成式人工智能模型简介;(2)活动 2:GAN 的构成;(3)活动 3:GAN 的应用;(4)活动 4:深度伪造;(5)活动 5:使用生成式模型进行创作。项目团队还设计了学习轨迹图(Learning Trajectory),方便教师了解学生对生成模型的技术构成、应用和伦理含义的理解,以在必要的时候给予引导。

为了更好地验证课程设计以及学习轨迹图的使用效果,项目团队以前后测的准实验方法面向 72 名学生(五年级至九年级)教授该课程,并在课后依据学生反馈对课程设计以及学习轨迹图的设计进行反思与修正。实践发现,学生课后测试正确率显著高于课前测试,学生对人工智能生成内容的辨别意识和能力有明显提升。然而,虽然学生们能够理解 GAN 如何通过生成器和鉴别器生成新的内容,但他们在理解输入数据集以及输入数据集如何关联到生成模型方面存在困难。项目团队认为学生们缺乏如何通过数据训练网络的基础人工智能知识^[22]。

3. “与人工智能共舞”课程

考虑到舞蹈、艺术、体育运动以及带有手势交互界面的电子游戏会激发学生兴趣,项目团队开发了一套基于交互式人工智能算法的模块化编程工具(包括 Teachable Machine 中的“手/身体/面部位置跟踪和表情检测”模块及“导入个性化图像和姿势识别模型”模块)和学习工具包,学生可以通过手势、姿态、面部位置与表情体验人工智能算法,设计交互式多媒体作品。项目团队还基于 TensorflowJS 等开源代码库开发了支持浏览器界面的编程平台^[23]。在该平台上,学生设计和构建他们感兴趣的人工智能项目,训练有监督机器学习模型,并思考交互式人工智能系统在日常生活中潜在的道德问题。课程为期五天,面向 21 名五年级至九年级的学生。

课后学生访谈结果显示,学生能够掌握和该交互式人工智能平台互动的方法,理解人机互动中数据表

示的重要性,能在一定程度上总结归纳人工智能姿态识别的原理及人机交互中的注意事项。针对交互式人工智能潜在伦理问题,学生表示可以通过修改人工智能算法规则以及调整数据集来减少人工智能算法偏见。

(二)“家校社”合作开展“数据行动主义”项目

作为人工智能伦理的重要议题之一,算法偏见被认为是社会偏见在算法信息社会的延伸^[24]。系统性种族歧视作为美国长期存在的社会偏见也延伸到了算法之中,导致潜在算法歧视。为了让非裔美国学生能意识到人工智能伦理问题的影响,同时让他们在人工智能及相关科技领域获得更多话语权,PRG 硕士研究生雷切尔·沃克(Raechel Walker)在2022年与当地社区组织合作,依托马萨诸塞州剑桥市的市长暑期青年就业计划项目,设计了为期六周的“数据行动主义”课程。除沃克外,软件工程师马修·泰勒(Matthew Taylor)和PRG博士后杜小雪(Xiaoxue Du)合作开发课程和设计跨学科体验项目,四名本科生通过MIT“本科生研究机会计划”参与了设计项目数据分析活动。

项目团队开发了一系列交互式编码和项目式活动(如表5所示),让学生重点了解人工智能系统中潜在的种族歧视,利用数据科学技能分析并揭露人工智能伦理问题对弱势群体的不利影响(如算法偏见等),探讨如何维护他们的权利,实现合乎伦理的机器学习,讨论如何设计人工智能算法并建立人工智能公平性的指标来实现种族平等。此外,该项目意在培养非裔美国学生对人工智能技术的批判性思维以及利用人工智能为社会作出贡献的信心。项目不要求学生具有编程基础,学生们在项目中学习使用文本编程(Python)可视化数据^[23]。项目开始前,PRG邀请了三名高中生进行预授课,根据其反馈迭代课程设计。在整个项目进程中,项目团队不断收集学生对活动的感受与建议,并根据学生的兴趣安排课程。该项目自2022年起已招收了两届学生^[25]。第一届是面向社会招募的11名非裔美国学生,并于2022年暑期开展了为期六周的线下课程。第二届学生则于2023年2月28日至4月22日(共10周)通过周末线下教学、周中混合式教学的方式参与活动^③。

表5 2022年“数据行动主义”项目活动概览

活动名称	活动内容	活动目标
可视化身边的数据	学生记录一周内与自己身份(性别、肤色等)相关的数据并进行可视化	学生意识到身边的数据可被用来分析;意识到身边不平等现象及其对生活的影响
数据存在歧视	学生使用“数据营养标签”标注未检测到的种族偏见的数据	学生意识到数据中潜存的种族歧视,知道并愿意采取行动反抗这一现象
赠人玫瑰	学生在老师的辅助下为低年级学生设计“数据行动主义”课程活动	学生意识到其可以对社会产生有意义的影响
启发性审计	学生通过刺绣、短文、绘画、舞蹈、时装等方式展示数据科学如何积极和/或消极地改变其生活	学生意识到尽管种族歧视存在,但自己可以达到很高的学术成就
用数据科学改变世界	学生通过数据挖掘、数据清理、探索性数据分析、可视化人工智能算法结果为受到人工智能算法歧视的非裔囚犯辩护	学生能够发现并指出社会不公平所需的批判性思维、编程技能和说服他人的能力

四、MIT 面向青少年的人工智能伦理教育实践对我国的启示

2022年底,由Open AI人工智能实验室发布的对话式大型语言模型ChatGPT实现了人工智能与人类流畅互动,其自动生成的文本在很多场景中以假乱真,人类已经进入智能社会新的发展阶段。智能技术的各种利弊正引发全球专家学者热议。在这样的背景下,人工智能伦理教育的重要性和迫切性更是日益凸显。自2018年以来,MIT在青少年人工智能伦理教育领域做出一些有益探索,我们从中可以得到以下

几点启示。

(一)高校可以同中小学合作开展青少年人工智能伦理教育的内容设计与教师培训

MIT不但在人工智能技术研发与应用方面位列世界高校前列,其在科技伦理研究和教育实践方面也在全球具有引领地位^[26],MIT媒体实验室近年来对青少年人工智能伦理教育的实践探索凸显了该校人工智能研发团队的社会责任担当,也体现了世界顶尖大学在智能社会中从事科研工作的价值引领,值得国内高校学习和借鉴。2018年,教育部发布了《高等学校人工智能创新行动计划》,其中明确指出高校是“科技

第一生产力、人才第一资源、创新第一动力的结合点”,在人工智能基础理论、关键技术研究及应用方面具有鲜明特色^[27]。因此,我国高校应尽快建立人工智能多层次教育体系以培养多样化的人工智能人才。同时,我国设有人工智能专业的高校可以发挥自身专业和师资优势,加强与中小学校的沟通和合作,在中小学生人工智能伦理教育方面有所作为,这对提升下一代人工智能素养有着重要的战略意义。人工智能伦理属于跨学科交叉领域^[28],面向青少年的人工智能伦理教育最好结合来自多个学科学者的意见和建议。

在具体操作层面,MIT 人工智能伦理教育项目开展模式值得我国借鉴。MIT 项目团队以学科教授为引领、研究生为主导,前期通过调研中小学一线教师对于人工智能及其伦理教育的“痛点”,有针对性地进行项目设计以及学习资源开发,随后以与人工智能教育特色学校合作或面向社会招募师生的方式开展项目。在此模式中,MIT 的角色在于将领域中新的理念引入课程设计,开发课程资源。此举既发挥了高校优势,又解决了一线教师课程资源中创新思维等综合素养和知识储备不足、时间精力不足等难题。而结合一线教师的实践经验再次优化课程设计,并将教学实践交给熟悉场景的中小学教师使得项目实施效果和效率最大化。

为进一步优化人工智能伦理教育项目实施效果,MIT 在项目开展前高校均对授课老师及助教进行培训。只有教师精准地掌握人工智能基础知识、伦理议题及教学设计要点,学生才能在其教导下正确理解人工智能及其伦理议题,在日常生活和学习中积极、负责任地使用人工智能技术。我国中小学各学科教师(尤其是信息科技教师)是推进青少年人工智能伦理教育的核心力量。因此,我国在开设人工智能伦理教育时同样需要重视高校教师与中小学教师的合作。高校可以通过对在职教师和师范生开展有关人工智能基础知识及其伦理议题的教师培训及举办相关研讨活动,更好地推动我国青少年人工智能伦理教育实践。

(二)中小学多学科教师可以尝试开展内嵌式青少年人工智能伦理教育的设计与实践

教育部新修订的“新课标”明确指出,中小學生要“正确应对人工智能对社会的影响,认识到人工智能对伦理与安全的挑战”。不过,在当前的教学实践中,中小学信息科技课程仍存在课时稀缺、教师有关人工智能伦理相关的知识和技能缺乏等问题^[29],因此,在绝大多数学校开展独立的人工智能伦理教育时机并

不成熟。也因此,内嵌式伦理教育实践更具可行性。内嵌式伦理课程最初由哈佛大学应用于高校计算机类学生的伦理教学中,指根据学生对一门非伦理类课程的需求,教学团队将伦理问题与该课程本身内容结合起来进行课程设计^[28]。总结 PRG 项目团队的实践经验可以发现,2022 年以来的项目均将人工智能伦理内嵌到计算机或人工智能课程中。因此,内嵌式伦理课程相比于专题式人工智能伦理课程可实施性更高。

此外,“新课标”突出跨学科内容的设计与实施,指出“要设立跨学科主题学习活动,加强学科间相互关联,带动课程综合化实施,强化实践性要求”,并为每个学段设计了跨学科主题活动。因此,中小学信息科技教师可以跨学科开展内嵌式人工智能伦理教育活动。例如,在第一学段跨学科主题活动“数字设备体验-信息安全小卫士”中,教师可以安排学生扮演信息安全小卫士等活动,运用信息科技、道德与法治、语文等多学科知识,加深学生对信息安全与个人隐私保护等人工智能伦理问题的理解和思考。

除了信息科技教师按照“新课标”设计跨学科主题学习活动外,其他学科教师也可以在开展跨学科活动时将人工智能伦理嵌入课程中。例如,在艺术课上,教师可以引入生成式人工智能技术,让学生在体验技术的过程中思考人工智能生成内容的影响(如版权所属问题等);在科学课上,教师可以利用图像识别等技术,让学生意识到人工智能对科学领域研究的影响以及其可能造成的决策偏差;在语言课上,教师可以介绍自然语言处理技术对于写作的影响,引导学生思考其在“人-机”互动中所扮演的角色并思考如何在日常合乎道德地使用该技术。其他学科教师日常授课时也可通过内嵌式伦理课程设计帮助青少年形成科学正确的伦理价值观。例如,历史课教师在讲授“工业革命”相关内容时,可以介绍工业革命给英国等国家带来的社会道德问题^[30],引导学生将工业技术社会影响的辩证思考迁移到对人工智能技术的思考中;生物课教师可以在介绍人体生理与健康的神系统时,适当讨论脑机接口技术及其带来的隐私保护等伦理难题^[31],鼓励学生自发交流讨论,引导其关注并探讨人工智能技术带来的伦理问题并形成正确的科技伦理观。

(三)“家校社”应携手共建多样态多场域本土化青少年人工智能伦理教育生态

青少年与人工智能技术互动的场景往往在校外,课堂上单一场景的教学并不能实现对广大青少年开展人工智能伦理教育,需要“家校社”三方共同努力,为青少年提供多样态(项目化学习、研讨会、工作坊

等)和多场域(家庭、学校、社区以及三方联动)的活动。MIT 媒体实验室不仅在学校教学场景中开展人工智能伦理教育的项目,还面向家庭场景和社区场景。具体而言,除了与初中合作开展项目外,项目团队在面向中小学生的“人工智能和数据隐私系列活动”中,设计了家庭场景中父母与青少年讨论人工智能伦理问题的活动指南;在“数据行动主义”活动中,与社区开展合作,依托社区活动项目开展人工智能伦理教育工作坊。MIT 媒体实验室通过多样态的项目活动促进“家校社”三方共同构建青少年人工智能伦理教育生态的做法值得借鉴。

2017 年,国务院印发《新一代人工智能发展规划》明确指出,“支持开展形式多样的人工智能科普活动……支持社会力量参与其中……面向公众开放人工智能展馆”^[32],体现了我国对“家校社”携手开展人工智能及其伦理科普教育的重视。我国中小学校可以进一步发挥家长的协同育人力量,鼓励更多青少年家长在家里与孩子多多交流,深入了解青少年使用人工智能技术的经历和他们对人工智能技术的态度,有意识地开展人工智能伦理问题的亲子讨论,引导其形成正确的人工智能技术伦理观和使用习惯。学校除了在日常授课中嵌入式开展人工智能伦理教学外,还可以在夏令营、冬令营以及研学活动中融入相关内容。社会一方面可以在科技场馆展览中增加人工智能体验活动并科普对应的伦理内容,另一方面可以和高校、中小学合作,在社区开展的活动中加入人工智能科普、体验及对其伦理问题的讨论。

伦理问题有“在地化”(Localization)的特点且具有强应用性^[9],MIT 的 PRG 项目团队多次实践同样表明,青少年能够更全面地分析日常生活中的人工智能技术及其伦理问题。因此,为了更好地在我国青少年一代普及人工智能知识和技能,整体提升其人工智能素养,“家校社”应携手设计适合我国国情的青少年人工智能伦理教育内容。“家校社”三方设计并开展青少年人工智能伦理教育时也应将日常生活情景相关联的社会现象等作为学习材料,鼓励青少年讨论、辩论与思考。具体而言,“家校社”开展人工智能伦理教育前,建议针对青少年人工智能伦理认知、人工智能使用等现状进行调研。根据调研结果,“家校社”三方可以选择更切合学生实际需求和偏好的课程主题、教学内容及学习活动,从而让人工智能伦理课程更加贴近其日常生活。我国青少年成长的社会文化背景以及所处人工智能生态环境与国外不同,唯有从我国文化视角下思考人工智能带来伦理问题,提出具有我国特色的人工智能伦理观乃至科技伦理观,形成具有我

国特色的青少年人工智能伦理教育体系,才能真正做好面向我国青少年的人工智能伦理教育。

【注释】

①《CSTA K-12 计算机科学标准》描述了一整套计算机科学核心学习目标,旨在为完整的计算机科学课程及其在 K-12 级别的实施奠定基础,该标准 2017 年由美国计算机科学教师学会(Computer Science Teachers Association,简称 CSTA)发布。CSTA 成立于 2004 年,旨在支持和鼓励计算机科学领域的教育,在计算机学科标准制定、教师培训上做了大量工作。

②自 2018 年起由美国国家科学基金会与卡耐基梅隆大学计算机科学学院资助,美国人工智能促进协会(Association for the Advancement of Artificial Intelligence,简称 AAAI)和 CSTA 联合发起的 K-12 人工智能(AI)计划(The Artificial Intelligence for K-12 Initiative,简称 AI4K12)。AI4K12 于 2019 年发布了《K-12 人工智能教学指南》(K-12 AI Guidelines),其中明确提出了五个有关人工智能的大概念,即感知、表示和推理、学习、自然交互以及社会影响。五个大概念均强调了人在人工智能中的重要作用,以引导对人工智能技术使用时的安全责任、歧视等伦理问题的探讨。

③实践结果具体信息以及第二届学生招募信息尚未公开。

【参考文献】

- [1]潘恩荣,曹先瑞.面向未来工程教育的人工智能伦理谱系[J].高等工程教育研究,2021(6):38-43+67.
- [2]张兆翔,张吉豫,谭铁牛.人工智能伦理问题的现状分析与对策[J].中国科学院院刊,2021(11):1270-1277.
- [3]黎常,金杨华.科技伦理视角下的人工智能研究[J].科研管理,2021(8):9-16.
- [4]刘肖肖,徐丽娟,刘慧晖.中国科技伦理治理体系的构建——中国科学院学部 2021'科技伦理研讨会会议综述[J].科学与社会,2021(4):138-141.
- [5]科技部.《新一代人工智能伦理规范》发布[EB/OL]. [2023-10-15]. https://www.safec.gov.cn/kjbgz/202109/t20210926_177063.html.
- [6]新华网.中国提交加强人工智能伦理治理立场文件[EB/OL]. [2022-11-16]. http://www.news.cn/world/2022-11/16/c_1129134620.htm.
- [7]教育部.义务教育信息科技课程标准(2022 年版)[EB/OL]. [2022-04-21]. <http://www.gov.cn/zhengce/zhengceku/2022-04/21/5686535/files/1b0c959dafc477dac723dc485095e49.pdf>.
- [8]张珊珊,杜晓敏,张安然.中小学开展人工智能教育的挑战、重点和策略[J].中国电化教育,2020(11):67-72+96.
- [9]任安波,叶斌.我国人工智能伦理教育的缺失及对策[J].科学与社会,2020(3):14-21.
- [10]古天龙.人工智能伦理及其课程教学[J].中国大学教学,2022(11):35-40.
- [11]何艺宁,商容轩,于游.高校人工智能课程与伦理道德教育融合探索[J].教育教学论坛,2019(41):144-145.
- [12]Saltz J, Skirpan M, Fiesler C, et al. Integrating ethics within machine learning courses[J]. ACM Transactions on Computing Education (TOCE), 2019(4): 1-26.
- [13]Kilhoffer Z, Zhou Z, Wang F, et al. "How technical do you get? I'm an English teacher": Teaching and Learning Cybersecurity and AI Ethics in High School[C]//2023 IEEE Symposium on Security and Privacy (SP). IEEE, 2023: 2032.

- [14] Lin X F, Wang Z, Zhou W, et al. Technological support to foster students' artificial intelligence ethics: An augmented reality - based contextualized dilemma discussion approach[J]. Computers & Education, 2023(201): 104813.
- [15] RAISE. RAISE program[EB/OL]. [2023 - 03 - 22]. <https://raise.mit.edu/>.
- [16] Ilana Pelzman - Kern. MIT Media Lab - AI + Ethics Curriculum for Middle School[EB/OL]. [2022 - 12 - 14]. <https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/>.
- [17] Payne B H. Can my algorithm be my opinion?: an AI + ethics curriculum for middle school students[D]. Massachusetts: Massachusetts Institute of Technology, 2020.
- [18] Ali S, Payne B H, Williams R, et al. Constructionism, ethics, and creativity: Developing primary and middle school artificial intelligence education[C]//International workshop on education in artificial intelligence k - 12 (eduai'19), 2019(2): 1 - 4.
- [19] Nguyen S. Cybersecurity Badge Day 2020: Girl Scouts of Eastern MA focuses on data privacy and tech policy[EB/OL]. [2023 - 10 - 23]. <https://www.media.mit.edu/posts/data-privacy-policy-to-practice-with-the-girl-scouts/>.
- [20] Williams R, Ali S, Devasia N, et al. AI + ethics curricula for middle school youth: Lessons learned from three project - based curricula[J]. International Journal of Artificial Intelligence in Education, 2023(2): 325 - 383.
- [21] Williams R. How to Train Your Robot: Project - Based AI and Ethics Education for Middle School Classrooms[C]. Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE'21), 2021:1382.
- [22] Ali S, DiPaola D, Breazeal C. What are GANs?: introducing generative adversarial networks to middle school students[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021(17): 15472 - 15479.
- [23] Jordan B, Devasia N, Hong J, et al. PoseBlocks: A toolkit for creating (and dancing) with AI[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021(17): 15551 - 15559.
- [24] 贾诗威, 闫慧. 算法偏见概念、哲理基础与后果的系统回顾[J]. 中国图书馆学报, 2022(6): 57 - 76.
- [25] Walker R, Sherif E, Breazeal C. Liberatory computing education for african american students[C]//2022 IEEE Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT). IEEE, 2022: 95 - 99.
- [26] 孙孟新. 国内外科技伦理研究的计量与比较[J]. 科学学研究, 2011(4): 481 - 486.
- [27] 教育部. 教育部关于印发《高等学校人工智能创新行动计划》的通知[EB/OL]. [2023 - 10 - 03]. http://www.moe.gov.cn/srcsite/A16/s7062/201804/t20180410_332722.html.
- [28] Grosz B J, Grant D G, Vredenburg K, et al. Embedded Ethics: integrating ethics across CS education[J]. Communications of the ACM, 2019(8): 54 - 61.
- [29] 柏宏权, 王姣阳. 中小学人工智能课程教师胜任力现状与对策研究[J]. 课程. 教材. 教法, 2020(12): 123 - 130.
- [30] 姜德福. 英国工业革命时期的道德问题——读恩格斯《英国工人阶级状况》[J]. 贵州社会科学, 2017(11): 57 - 61.
- [31] 肖峰. 脑机接口技术的伦理难题与应循原则[J]. 中州学刊, 2022(7): 95 - 102.
- [32] 国务院. 国务院关于印发新一代人工智能发展规划的通知[EB/OL]. [2023 - 09 - 20]. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.

Exploration and Inspiration of Artificial Intelligence Ethics Education for Teenagers ——A Case Study of MIT

LI Yan, ZHU Yumeng, FAN Xiaoyu
(Zhejiang University, Hangzhou Zhejiang 310058)

Abstract: As the application of artificial intelligence continues to grow, the issue of Artificial Intelligence (AI) ethics has become increasingly prominent. The Chinese Compulsory Education Information Technology Curriculum(2022 edition) clearly states that AI ethics education should be provided to primary and secondary school students. Given that AI ethics education for teenagers is still in its infancy, the introduction of Massachusetts Institute of Technology (MIT)'s exploration in this area since 2018 can provide insights into the teaching practice of AI ethics education in primary and secondary schools in China. MIT's AI ethics education programs for teenagers covered three school levels: elementary, middle and high school, with a focus on middle and high school, and involves three domains: family, school and society, in both thematic and embedded forms. The programs included offline STEAM elective courses in collaboration with elementary and middle schools, online/offline summer camps, and hybrid workshops for recruiting students from the community. By summarizing MIT's experience, this article provides three suggestions for the further development of AI ethics education for Chinese teenagers: (1) colleges and universities can collaborate with K - 12 schools to design curriculum and training teachers for AI ethics education, (2) K - 12 school teachers from various subjects can attempt to undertake embedded ethical education design and practice for youth AI ethics education, and (3) Home, school, and society should collaborate to establish a diverse and multiple fields - related localized AI ethics education for young people.

Key words: Artificial Intelligence Education; Artificial Intelligence Ethics Education; Teenagers; MIT