教育人工智能伦理研究现状和趋势
—基于系统性文献分析研究

郭昱含
北京邮电大学　北京　100000

摘　要：近些年来，随着人工智能算法突飞猛进地发展，社会逐步进入“智能时代”。各种智能技术在赋能教育领域的同时，也带来了严峻的伦理挑战。搭载了人工智能算法的计算机，逐步从被动工具变成教学情境下人类的代理者，这引发了社会各界对人工智能运用于教育问题的伦理的思考和担忧，我们急需建立新的伦理范式，将人类社会的伦理规范延伸到教育领域下的智能个体。本文回顾了近八年来的相关研究，依据现有的EEAI框架梳理当前针对教育人工智能伦理的研究领域及内容，以及相应的伦理问题解决进路。在总结已有研究的基础之上，提出我国教育人工智能伦理治理的新方案。

关键词：人工智能伦理；伦理风险；综述

引言

近些年来，人工智能（artificial intelligence，AI）与大数据技术的发展带领人们进入了全新的智能时代[1]。人工智能技术在给生活带来的便利的同时，也改变了居民的生活方式，作为一种全新的技术，应用于教育具有天然的合理性，人们对新技术赋能教育是充满期待的。因此，人工智能伦理研究是人工智能时代的必然产物，它既包括对技术本身的研究，也包括在符合人类价值的前提下对人、机和环境之间的关系研究[2]。

不同于工业生产领域，教育需要考虑到被教育者的人格塑造与培养，新技术的开发与应用必须要考虑各种价值观念与伦理问题[3]。虽然AI技术极大地推动了教育数据采集运用，提升了教学效果，为课堂授课带来了诸多便利，但作为一种新技术也孕育着我们无法预知的风险，例如数据误用、数据标准欠缺、数据过度依赖、数据监管不严、数据应用失范以及数据隐私泄露而引发的数据伦理风险[4]。

本文的目的是借助系统性文献综述方法，厘清现行国内外教育领域人工智能伦理研究的发展脉络、现状和趋势，以期对未来的理论研究和实践提供借鉴。

一、核心概念和理论框架

（一）核心概念

1.人工智能

从概念上看，（Artificial Intelligence，人工智能简称AI）由于其具有较强的跨学科性质，所以其概念也有较强的复杂性。源于视角的差异，人类学家、生物学家、语言学家、哲学家、心理学家和神经科学家都对人工智能发展作出了贡献，每个群体都带来了自己的观点和术语，这无疑导致人工智能的概念复杂且多样。

教育研究者普遍认为人工智能的术语起源于1956年生物学美国达特茅斯会议John McCarthy等人提出的“使用机器模拟人类智能”概念[5]，McCarthy教授将人工智能视为一种制造智能机器，特别是智能计算机程序的科学和工程，且与使用计算机来理解人类智力的类似任务有关；我国工程院院士、国际欧亚科学院院士李德毅先生认为，目前的人工智能与智能科学技术是“同义词”，是在人脑认知启发下发展的一种综合学科[6]，并可以“脱离人类意识存在”，成为人类智能的“体外延伸”。

2.教育人工智能

教育人工智能（EAI）是国内教育研究者使用的专业词汇，首次被闫志明[7]等总结、定义为“人工智能与学习科学的融合”，目的是通过人工智能技术解析学习过程，利用人工智能的相关工具探究学习的发生规律和作用机制。徐晔[8]解释EAI是人工智能与教育融合发展的“高级阶段”。是教育领域人工智能发展的“应然形态”。EAI通过人工智能技术理解和窥探学习过程。尹璐[9]等人从哲学角度切入，在本体论的视域下，从施教者、受教者与第三方参与者的角度出发，阐述清楚教育人工智能的“双主体”意蕴，并在认识论的视域下，明确人工智

能作为物体的“工具理性”强化人的主体意识。

3.教育人工智能伦理

伦理是指社会生产生活中处理人与人、人与社会相互关系时应遵循的道德观念和行为准则。教育伦理既包括在师生关系、家校关系、人才培养中存在的一般社会道德，也包括与教育内涵相符合的社会价值属性[10]。教育首先是对人类文明的代际传承和社会传播，即包含人才培养又包含文化传承，所谓人才也即能学习既有专门知识并能持续创新。文明传承是教育最基本的价值体现，也应成为对一切教育行为进行善恶评判的衡量标尺。概言之，所谓教育伦理，指的是文明传承教育工作需要的思想观念和行为规范。

当人工智能被运用于教育时，它强大的数据整合与分析能力，同时也会引发一系列新的伦理问题，例如：可能会给用户的私人数据和隐私保护带来潜在的风险；模糊教师、学生和智能导师等角色之间的界限等。因此，教育人工智能的设计、开发和教学实践应用都需要对其伦理问题的复杂性有充分的认识，并认真思考应该遵循何种伦理原则[11]。

（二）理论框架

只有明确教育人工智能运用时可能遇到的伦理问题，这样它们才能真正做到智能技术赋能教育，尤其当使用人工智能驱动的系统对未成年人进行教学时，必须要考虑到人工智能技术对幼儿的影响，这迫切需要一个理论框架对问题进行分析与整合。目前为止，有以下的教育人工智能伦理分析框架。

2019年4月8日欧盟正式提出《可信赖的人工智能伦理准则》，确定人工智能需要具备三个必要条件：合法性、伦理性和稳健性，以及明确人工智能伦理框架的七个组成部分为“人的能动性与监督”，“技术稳健性与安全”，“隐私与数据管理”，“社会与环境福祉”“多样性、非歧视性与公平性”“透明性”和“问责制度”等七个方面。

国内已有的研究者，尝试从教育人工智能改变教育群体网络体系[12]入手，将教育人工智能伦理问题分为人工智能教育场下，人与人交互，以及人与人工智能交互下产生的伦理问题。同时明确，教育人工智能的伦理治理应从习俗迁移、规范构建、法律约束等方面着手。而人工智能时代的真正到来，一定包括多层次的人机协作[13]，这不仅包括人类个体与智能技术的互动、人类群体与群体智能技术之间技术媒介干预下的人际关联、主权国家的智能治理以国际智能技术合作与竞争、人机协作与环境和生态系统的相互制约等；苗逢春[14]通过对联合国教科文组织的《人工智能伦理问题建议书》进行教育解读，从机器决策与人文实践互动的维度界定教育人工智能伦理问题分析框架，剖析基于数据和算法的预测和决策引发的典型伦理问题。在此框架下，辨析教育人工智能伦理问题的主要表现形式和教育在培养人工智能伦理观念中的作用。也有研究者[15]以责任伦理为视角，提出教育人工智能风险治理的责任伦理框架，其涉及角色责任、契约责任、前瞻责任、关护责任四个维度。指出教育人工智能风险治理的责任伦理困境表现。

上述框架都是从单一，或是少数角度切入分析人工智能伦理问题。Erica[16]等人基于公民权利、伦理原则和在人工智能的世界中学习三个角度，设计了道德，教育与人工智能（EEAI）框架，总结出人工智能运用于教育需要的五大伦理支柱。为教师、学校领导和政策制定者提供了一种全面思考教育人工智能伦理问题的框架。Zhou[17]等人基于此框架进一步结合K–12教育问题的特点，制定了一份设计指南，以帮助研究人员和课程设计者规避潜在的伦理风险。Joshi[18]等人基于该框架分析了下一代教育人工智能在帮助教育者提高发展中国家教育的公平性和同时也会面临哪些潜在的问题。Popkhadze[19]等人基于该框架，探讨了教育人工智能如何在推动高等教育发展的同时，会在不警惕的情况下产生哪些不利影响。

在EEAI框架中，教育人工智能在使用中的可能遇到的伦理问题分为三个大的类别，分别是公民权利、伦理原则、学习，在这三大类问题之上，Erica等人总结出面向大众的人工智能设计、实施和治理的五大维度。

1.公民权利

教育人工智能的使用有可能会影响到公民权利，进而引发伦理问题。具体表现为，智能算法的自适应特性，会在潜移默化中收集与存储用户数据[20]，当用户无法参与并决定自身的数据的去向时，会导致无意识下的个人数据泄漏，侵犯个人隐私权。同样，算法本身的“黑箱”特性，导致设计者也无法明确决策细节，更谈不上对决策的合理评估与监管[21]，由此容易引发算法偏见，导致部分使用者受歧视，丧失了被平等对待的权利。并且，智能系统的自动决策与内在机理的不明确以及教育领域内人工智能法律规章的缺乏[22]，会在侵权问题发生后，难以在智能系统主体之间进行问责，从而导致公民

权利的进一步损害。综合以上问题，作者提出赋能原则，要求学习者能够了解自身权利，并参与法律法规的制定。

2.伦理原则

秉持诚信、正义、仁慈和尊重的当代道德原则是合乎道德和安全地使用人工智能驱动系统的关键。

3.学习

当学习者在使用教育人工智能学习时，也有可能出现相应的伦理问题。例如，由于人工智能算法本身的缺陷，以及使用者对人工智能算法了解的缺乏，当被用于教育时，会出现难以预估的后果。同样，当学习者在教育人工智能影响下进行学习时，也会出现相应的问题。

4.面向大众的人工智能设计、实施和治理

基于上述三个问题，Erica等人提出了五大伦理治理维度。

意识：意识的伦理维度要求保证使用教育人工智能的个人或团体的知情权，在使用人工智能技术前给予告知，以便他们能够采取对应的行动和决策。

可解释：可解释的伦理维度，包括两个方面。首先，是针对教育人工智能系统的使用者继续普及人工智能的知识，并以合理的方式解释有关智能技术应用的问题。其次，针对智能系统的设计者和使用者，应保证他们能够清楚地解释他们为什么使用人工智能驱动的系统，它打算做什么和实际做什么（包括是否出现与歧视和偏见相关的意外后果），系统如何做出决定，以及它的好处和风险。当人工智能系统造成伤害时，相关的责任人必须公开解释这是如何发生的以及他们将如何应对，不仅是对事件，而且是为了该技术在未来的继续使用。

公平：公平的伦理维度，要求避免由于智能算法在运用于教育时，而出现的算法歧视。在教育之外，人工智能的使用导致出现了许多偏见案例，这些算法偏见往往包含着性别歧视、种族主义和其他形式的歧视[23]。

透明：透明的伦理维度，要求系统设计者在智能系统用于教育之前，系统的设计者对于教育人工智能系统有充分的认识。人工智能通常被描述为一种不透明的技术。它通常以不可见的方式注入到计算系统中，从而影响我们的互动、决策、情绪和自我意识[24]。

问责：问责的伦理维度，要求相关的政策制定者以推进问责制度与法律法规的方式，来进行人工智能治理，明确智能算法的审查与监管标准，出台相关政策文件，设立风险评估环节，力求将使用AI需要承担的法律义务与社会、文化和经济责任相互联系起来。

二、研究设计

（一）研究问题

为呈现面向教育的人工智能伦理研究现状，本研究确定了以下的研究问题：

1）涉及到教育人工智能伦理的问题研究集中在哪些主题，有哪些研究趋势？

2）涉及到教育人工智能伦理的问题研究，还有哪些领域关注度较少？

3）涉及到教育人工智能伦理的问题研究，主要是来自于哪些领域？

本研究采用系统性文献综述方法。系统性文献综述旨在以明确的研究问题、全面的检索策略、清晰的文献纳入标准与综合的数据分析，来得出可信的研究结果，可减少由传统文献综述方法带来的研究偏倚[25]。

（二）研究样本获取以及筛选

1.文献检索策略

本文基于Web of Science（WOS）和CNKI数据库。在CNKI数据库中，以“人工智能伦理”，“教育伦理”，“教育人工智能”作为关键词进行搜索。在Web of Science数据库选择以人工智能伦理（AI ethics）、教育伦理（education enthics）、人工智能伦理（Artific-ial Intelligence enthics）等为关键词进行搜索。

2.文献筛选标准

为保障文献的质量，只检索web of science核心期刊与CSSCI核心期刊，选取2016—2023年（共8年）期间发表在核心期刊期刊上有关教育人工智能伦理研究的495篇文献。

为精确掌握教育人工智能伦理问题研究现状，本研究根据研究问题对检索得到的文献制定了相应的文献纳入和排除标准。具体按“全文可获取”、“发表时间处于2016-2023”、“全文篇幅至少3页”、“研究对象为教育人工智能伦理问题”、“研究问题包含确切的研究问题、研究方法与研究结论”、“中文期刊来自于CSSCI”、“英文期刊来自SSCI”标准进行筛选，其中前3条在进一步提升文献样本筛选的精确性；后四条分别对研究对象、研究内容、研究目标、研究过程与研究层次做限定，以便更好地聚焦研究问题。

3.文献筛选过程

本研究遵循系统性文献综述方法，通过检索、筛选、合格和纳入四个阶段对检索到的教育人工智能伦理文献进行筛选。最终纳入有效文献中文文献35篇，英文文献

28篇。

（三）文献编码策略

为了系统性地研究教育人工智能伦理领域的研究热点与趋势，本文依据EEAI模型，对检索到的文献进行分类编码。

此处对文献编码表格做进一步解释：在EEAI模型公民权利部分，赋能原则的提出是作者对解决缺乏参与、问责困难以及法律不健全等三类问题的呼吁，因此不单独作为研究领域的编码。而公民权利下的问责与合法性要素，都是针对侵权问题发生后责任分配等问题，因此编码合并；面向大众的人工智能设计、实施与治理是基于三类伦理问题总结出的治理框架，因此不参与编码。

因此，最终编码维度分别为公民权利、伦理原则、学习三大类，“公民权利”指教育人工智能的过度使用有可能会侵犯到哪些公民权利；“伦理原则”指教育人工智能的使用，“学习”指应该秉承怎样的伦理原则；教育人工智能的使用中会对学习者提出了哪些要求，会有哪些负面影响？

三、研究主题与发展趋势

（一）公民权利

通过对已有文献进行分析，我们发现，在所有检索到的文献中。52%的文献未针对公民权利进行讨论，31%的文献讨论了一个公民权利要素，17%的文献针对了两个及以上的公民权利要素展开讨论。这说明，目前的学界已有相当一部分研究者，注意到了教育人工智能会造成公民权利潜在伦理问题，并且针对公民权利的某些方面展开进行论述。但是论述的内容往往只涉及到公民权利中的一到两个要素。且“参与”以及“问责与法律框架”这两个类别的研究数量，要明显高于“非歧视”研究类别。笔者认为，由于人工智能算法本身的自动化和不可解释等特性，因此学术界比较关注智能系统在学习者未察觉的情况下，对其可能造成的潜在伤害。并且，试图探讨如何明确该类事件发生后的问责机制，以及相关的法律法规。

根据已有的文献分析框架，公民权利领域的研究文献主要被分为以下三类。

第一类文献即“参与”类别。该主题的文献，关注由于使用者缺乏对AI系统中使用数据环节的参与，从而导致的个人隐私权被侵害等。对于个人隐私的侵害，主要体现在未经许可的数据存储与数据使用。

在数据存储方面，冯锐[30]等人注意到，学习者使用系统时的产生的行为数据有可能长久地存储在某些服务器上，如果不明确该类数据的有效期限和使用范围，将意味着数据产生后的数十年内，该数据被使用或滥用的可能性持续存在。

在数据使用方面，候浩翔[20]等人注意到，由于个人数据信息的挖掘处理与关联性分析成为创造学习价值的重要形式，许多企业聚拢的学习者相关信息在深层次加工后，超出了最初的数据收集形态，用于无法预知的盈利需要乃至其它非法目的。

但是也有学者发现，没有海量的数据支持，智能系统无法洞悉教育规律[13]，目前便捷个性化的学习推送服务往往要以学习者的泄露隐私或知情同意权的缺位作为代价。基于此类情况，有研究者预测[21]，如果不提高学习主体对数据管理的知情与参与程度，并对该类问题进行伦理制约，这不仅会导致严重的个人隐私泄漏问题，还会导致相关的教育数据被第三方任意使用或篡改，引发更为复杂的数据安全伦理问题，从而使社会缺乏新技术运用于教育的信任[35]，最终导致在教育人工智能应用的可持续发展失去可能。

为避免上述情况的发生，已有李晓岩提出了人工智能教育应用的尊重原则[32]，强调数据在收集以及使用过程中，应让学生以及教师主体参与其中，行使决定权，来绝对这些数据能否被收集，在何种范围内使用。为了平衡个人隐私泄漏问题与智能教育系统对数据渴求之间的冲突，刘三女牙[33]建议任何公共教育数据的公开与共享，都要尽量避免涉及学习主体个人的细致信息，且任何学习对象对在享受学习服务过程中产生、搜集、存储与自身有关的数据都要享有知情权，并且相关数据需要得到有效的保护，在获得学习者授权的情况下方可合理使用。

第二类文献即“问责与法律”类别。该主题的文献，关注教育人工智能损害部分权益时，各个责任主体之间的责任分配以及相关的法律规章缺乏的问题。随着人工智能逐步渗入教育领域，各国相继颁布相关法律法规以及指导政策。

2002年，美国通过《教育科学改革法》，试图在高等教育领域率先推行人工智能技术，其他国家也纷纷加大对教育信息化的建设与推进力度。2015年以来，人工智能在国际教育领域的应用逐渐增多，我国也高度重视人工智能在各领域的应用。国务院于2017年7月印发《新一代人工智能发展规划》，部署人工智能背景下

创新型国家和世界科技强国建设。2019年，国家新一代人工智能治理专业委员会发布《新一代人工智能治理原则——发展负责任的人工智能》，为人工智能治理与风险防范提出科学指导，致力于提升人工智能的安全性与可靠性。2021年11月，联合国教科文组织发布了《人工智能伦理问题建议书》。因此，也引发了一系列对政策相关政策的解读[14]，以及对问责机制的研究。

但是高山冰[34]等人分析了上述规章制度，以及法律法规后发现，虽然在人工智能伦理准则设计、智能应用整治等方面取得了一些成果，但与教育人工智能相关的立法工作的仍旧无法应对教育人工智能在使用时会出现的问题。

对于产生这一情况原因，有研究者[35]认为，是教育治理体系和教育治理能力现代化程度不够，现有的教育法律模式以及伦理规范并未跟上快速发展的技术，造成智能教育治理缺乏有效的约束与引导。比如深度学习技术在人工智能领域的广泛使用[20]，其自适应和自主决策的特征，极有可能以人类无法认知地方式，做出危害性行为，这给问责制度以及法律法规的确立，带来了极大的困难。所以有研究者[12]认为，教育人工智能的自动决策和复杂特性，使人们难以了解其内在机理，是导致难以确立问责机制的主要原因。

赵磊磊[18]结合了上述观点，认为由于人工智能技术本身作为技术产物的不确定性，以及目前各个教育责任主体间的利益和权责不清，共同导致了问责困难。

如果想要解决上述情况，明确在教育场景下智能算法的问责机制并确立相关的法律，首先需要理解智能教育系统内部的运行机制[37]，以技术和潜在风险的可解释性为基础，对伦理问题进行治理。但仍有一部分研究者认为，解决伦理问题的关键在于建立相关政策以约束技术，避免技术异化[38]，消除人工智能系统可能带来的某些不恰当的价值判断，以免在技术快速演进、法律法规相对滞后的当下引发一系列社会问题和伦理挑战。

第三类文献即“非歧视”类别。该主题的文献，关注由于人工智能算法偏见，导致部分使用者受歧视，侵犯了作为个体被平等对待的权利。已有研究者[34]认为，在教育领域运用智能系统时，由于算法本身的复杂性以及数据的中立性、客观性难以保证，极有可能导致算法偏见的发生。例如2016年微软聊天机器人Tay上线不久便学会了辱骂用户[30]，发布带有种族歧视和性别歧视的言论。学习者在与人工智能交互过程中，自然语言处理技术和深度学习技术使人工智能习得了部分使用者的带有负面、消极情绪的语言表达模式。

（二）伦理原则

已有相当数量的研究探讨了人工智能伦理原则，要求从有益、安全、可释、公平、稳健等方面打造合乎伦理的人工智能[39]，做到透明可释、公平公正、科技向善、责任担当、保护隐私等，充分体现了作为主体的人的发展需求。因此，大部分研究认为，应从教育伦理学视角研究教育人工智能，从“人”的角度出发，理解教育人工智能伦理的本质，辨明教育人工智能开发与应用过程中的关系与行为。李晓岩[32]认为，教育人工智能伦理原则应从“人”的角度出发，规约人的行为，于是针对教育管理者、教师、学习者等三个教学活动中不同的角色，提出了权利规约原则、和谐共生原则以及形塑自我原则，并且呼吁研究者应使用教育伦理切入研究，而非技术伦理。

但是也有一部分学者认为，教育人工智能伦理是一种技术伦理，可以参考人工智能伦理原则。邓国民等人认为，一系列新的伦理问题是由于教育人工智能的强大数据整合与分析能力而导致的，提出福祉，是非善恶，公平正义，人权和尊严，自由自治，责任和问责等原则。

同样，鉴于人工智能伦理交叉学科的特性，相近领域的研究成果也可被用于人工智能伦理治理。通过借鉴计算机和信息伦理学的PAPA的道德准则，即隐私权（Privacy）、准确（Accuracy）、所有权（Property）、易获得性（Acces-sibility）[41]，杜静[42]等人基于对人工智能伦理的相关要素分析，立足人工智能在教育领域的应用现状，从人机共存角度，抽取与学生发展密切相关的要素，涉及明确责任主体、保护人类隐私、不偏见不歧视、决策透明化、保护人类利益不受侵害、提前预警危险行为、系统稳定可控等多个方面，最终归纳出面向智能教育的人工智能伦理模型，即问责原则、隐私原则、平等原则、透明原则、不伤害原则、身份认同原则、预警原则、稳定原则，概括为“APETHICS”模型。

（三）学习

根据已有文献分析结果，尽管人工智能技术在教育领域得到了广泛的应用，但涉及学习维度的研究中，关于“人工智能算法”的文献数量较“在人工智能世界中成长”和“运用人工智能”相比较少，数量为其他两类的一半，笔者认为，由于人工智能算法本身具有高度的技术性和复杂性，这种跨学科的计算机复杂性可能导致非技术学者对算法的研究难度加大，而更多的研究倾向

于关注如何使用人工智能工具来促进学习或教育；且教育领域的焦点往往更关注人工智能在课堂教学、学生成绩提升等实际应用中的效果，而非其背后的技术实现，忽视了对“人工智能算法”本身对教育伦理的深入探讨。此外，由于人工智能算法的设计和实现主要是由技术公司和专业技术团队完成的，这些技术大多处于“黑箱”状态，一般的教育研究人员难以获取详细的算法信息。这种技术壁垒可能也是如此限制此类研究数量的一个重要原因。技术的封闭性和布局性使得学术界对人工智能算法的伦理分析在教育领域内的讨论尚未成熟。

老师的教与学生的学的本质，是一种培养人的社会活动，而且，这种培养过程是一种学生与教师之间的交互过程，如果学生在学习的过程中过度依赖和人工智能的交互，会导致学生的学习能力下降，最后导致学习能力丧失。因为，人工智能具有强大的数据获取和分析能力，当这一技术被运用于教育时，能根据学生学习的风格与薄弱环节精准推送学习资源[43]。但若学生长期接受教育人工智能的数据推送，学生将有可能形成“智能”依赖，丧失主动探索新知识的好奇心，丧失对自己学习薄弱环节查缺补漏的能力。也使学生丧失获取信息的自主权，剥夺了很多思维训练和学习体验的机会，改变他们的大脑结构和认知能力，尤其是在学习者大脑发育的关键期过度使用认知技术，将可能带来不可逆转的严重后果，使其行为和思维产生惰性[44]，久而久之，学生将不能作为学习的主体，自主选择自己的发展方向，而是仅仅作为推荐算法下的接受者，这对他们发展更高级的认知技能和创新能力是非常不利的[45]。

教学也是一个注重揭示教育现象或教育行为之间的因果关系。当前以深度神经网络为代表的智能算法普遍存在“黑箱”问题，虽然在多个教育场景或任务上都构建了精准、高效的机器学习模型，但却难以理解或解释其工作机理，无法进一步针对不同学习者的学习特点进行和错因追溯、归因分析，这将阻碍学生的个人发展与成长[46]。学生在与人工智能的交往中，由于缺乏与人的沟通与交流[47]，难以“学会生存、学会求知、学会做事、学会合作”。教育本身的人文特征将无法凸显。若应用人工智能成为惯性依赖甚至生存保障，其中可能隐含的人性反叛与人文关怀的迷失必然阻碍教育进步。当精准的预测服务远远超出人脑的控制水平时，学生及其家长很有可能不由自主地屈从于大数据勾勒出的职业发展轨迹[48]。而一个学生成长的过程，其性格、能力、价值观、道德品质等都会随着时间推移而不断地变化、发展和成熟，但过往数据却始终不变，过去学习生涯中的负面信息，有可能影响学生的改过自新与长远发展。并且，基于教育大数据的数字化记录不同于人脑的记忆[49]，不会有自身淡忘和外界干扰等问题，这就意味着，在一定程度上，承载学习个体种种学习行为所记录的数据，会在某些时刻被用于模型训练，或许起初施行此类操作的初衷是为学习者制订出效果更佳的个性化学习路径，但过去的种种数据却会为学习主体从此烙上具有某种暗指含义的固化标签，而使主体失去更多的自由选择。

最后，人工智能算法评价学生的方式，往往是通过学生掌握知识的程度，长此以往教学的过程蜕变为知识传输过程，学生变成知识的容器，衡量学生的标准也就变为学生掌握知识的多寡。这一问题的深层原因在于技术对教育的侵蚀[50]，人工智能教育应用面向的是教育信息的深度处理，教育的“唯技术论”可能加剧当前标准化教育的泛滥。各种智能工具造成的碎片化知识也会导致学生的浅表学习，削弱学生的创造力和想象力。总之，人工智能的情感盲区会造成学生社会情感学习的缺失，进而使人成为一个单向度的人，极大地影响学生的个人发展。

四、研究局限性及结论

既有研究也从前瞻性视角，指出目前教育人工智能伦理问题路径主要集中在当前教育人工智能伦理建设存在缺乏精准的规范与指引、人文关怀与价值引领缺位、技术自身存在局限、学习资源建设粗放等难题[51]。本文通过系统梳理“教育人工智能伦理应用”的中外学术文献，形成了教育人工智能应用伦理问题的风险、经验借鉴及未来路向等核心研究议题。总而言之，既有研究囊括了教育人工智能伦理研究的基础理论问题，初步搭建起了该领域研究的问题框架，为该领域的理论发展奠定了良好基础。

然而，既有研究还存在一些亟待补强的薄弱环节。

一是研究问题较零散，未形成集群化的问题体系。既有研究选取的问题视角都较零散，未在把握该领域国内外前沿研究议题基础上，围绕教育人工智能伦理研究的几个核心问题，形成一个问题间有紧密逻辑关联、问题有理论嵌入性的集群化问题体系。

二是研究方法单一，跨学科研究欠缺。既有研究较多采用定性分析方法，从理论角度剖析教育人工智能伦理的各项问题，极少文献采用规范的量化研究方法对具

体问题作实证研究。这反映出教育人工智能伦理研究在方法和工具上的单一化和初阶性。此外，利用跨学科理论、方法与视角的研究缺乏，既有研究的跨学科性还很薄弱，对于教育人工智能伦理问题，往往仅局限在单一的学科视角之内。

三是基础理论研究尚不深入。既有研究尚未深入分析教育人工智能伦理的基础理论。这一方面体现在核心概念的界定和辨析不足，既有研究多是技术哲学视角下的“人工智能伦理”概念，未对教育情景下的概念内涵、外延和特质性作深入讨论。另一方面，既有研究反映出理论基础的混用与误用严重，现有文献的理论假设、核心命题和观点往往不足以支撑教育人工智能伦理问题的推演与论证。这显示出，现有研究并未准确把握教育人工智能伦理研究的理论基础。

为破解既有研究的薄弱环节，本文认为教育人工智能应用的未来研究可进一步突破的方向为：

第一，构建一个国际化和集群化的问题框架。持续系统梳理国内外该领域前沿研究议题，从中挖掘该领域的重大核心问题，未来研究可集中围绕这些重大理论问题展开深入研究，避免研究问题的过度零散和细微化。在廓清重大理论问题的基础上，还应进一步扩宽该领域研究的问题面向，包括教育人工智能伦理的价值观规约问题、教育人工智能伦理规范指南的编写、教育人工智能伦理建设原则、编制教育人工智能伦理规范指南、提升教师智能教育素养水平、以智能技术反向赋能伦理建设，以及建设公益性学习资源等策略来规避风险等的议题依旧有待开掘。

第二，促进研究方法的多元化和跨学科性。未来研究可尝试采用案例研究、焦点小组访谈法、社会网络分析、扎根理论等规范质性研究方法，或问卷调查和实验研究等定量研究法，从政治学、哲学、法学、教育学、管理学、信息工程学、计算机科学和数学等多个学科，结合不同学科的视角，开展跨学科的研究，以便更全面地理解和解决教育人工智能伦理面临的复杂问题。这种多元化的方法论不仅能够丰富研究的深度和广度，还能为实际应用提供更加切实的指导。此外，研究者还应加强与实践者的合作，通过实地调研和实践反馈，不断修正和完善理论框架，以确保研究成果能够在教育实践中真正落地。

综上所述，教育人工智能伦理的研究应当以更加系统化、跨学科的视角为导向，推动理论与实践的结合，从而为教育领域的可持续发展奠定坚实的伦理基础。

结语

人工智能时代机遇丛生，也充满挑战。著名物理学家霍金曾表示：“人工智能发展到极致时，我们将面临着人类历史上的最好或者最坏的事情。”[52]教育人工智能为教育事业的发展作出了积极的贡献，但也给人类带来了严峻且迫切的挑战。如何最大程度地规避教育人工智能的伦理风险，稳步推进教育高质量发展，是值得深入探讨的问题。针对当前伦理风险频发的情况，本研究对教育人工智能伦理研究领域进行使用文献综述法认知现状调研，剖析了教育人工智能伦理的几个核心问题，并提出了具有前瞻性和针对性的建议。因此，国家相关部门和高校专家学者应重视该领域的研究，不断加快发展其风险监测技术，改善智能伦理认知现状及伦理建设难题，从而提出更全面的对策建议。但无论如何，人类都应充分发挥其主观能动性，直面可能产生的伦理风险并将伦理规范嵌入教育人工智能的全生命周期中，不断夯实智能教育发展和智慧社会建设的根基。

引用文献

[1]赵磊磊，张黎，王靖.智能时代教育数据伦理风险：典型表征与治理路径[J/OL].中国远程教育，2022（03 vo No.566）：17-25+77. DOI：10.13541/j.cnki.chinade.2022.03.004.

[2]刘伟，赵路.对人工智能若干伦理问题的思考[J].科学与社会，2018，8（1）：40-48.

[3]李子运.人工智能赋能教育的伦理思考[J].中国电化教育，2021（11）：39-45.

[4]刘三女牙，杨宗凯，李卿.教育数据伦理：大数据时代教育的新挑战[J].教育研究，2017，38（04）：15-20.

[5]徐晔.从“人工智能教育”走向“教育人工智能”的路径探究[J].中国电化教育，2018（12）：81-87.

[6]李德毅.AI——人类社会发展的加速器[J].智能系统学报，2017（05 vo 12）：583-589.

[7]孟翀，王以宁.教育领域中的人工智能：概念辨析、应用隐忧与解决途径[J/OL].现代远距离教育，2021（02）：62-69. DOI：10.13927/j.cnki.yuan.20210316.001.

[8]徐晔.从“人工智能教育”走向“教育人工智能”的路径探究[J].中国电化教育，2018（12）：81-87.

[9]尹璐，安维复，刘进.教育人工智能的哲学意蕴

[J]. 高教探索，2021（05）：39-45.

[10] 张永波. 智慧教育伦理观的建构机理研究 [J]. 中国电化教育，2020（03）：49-55+92.

[11] 邓国民，李梅. 教育人工智能伦理问题与伦理原则探讨 [J/OL]. 电化教育研究，2020，41（06）：39-45. DOI：10.13811/j.cnki.eer.2020.06.006.

[12] 张立国，刘晓琳，常家硕. 人工智能教育伦理问题及其规约 [J/OL]. 电化教育研究，2021（08 vo v.42；No.340）：5-11. DOI：10.13811/j.cnki.eer.2021.08.001.

[13] 苗逢春. 教育人工智能伦理的解析与治理——《人工智能伦理问题建议书》的教育解读 [J]. 中国电化教育，2022（06）：22-36.

[14] 苗逢春. 教育人工智能伦理的解析与治理——《人工智能伦理问题建议书》的教育解读 [J]. 中国电化教育，2022（06）：22-36.

[15] 赵磊磊，吴小凡，赵可云. 责任伦理：教育人工智能风险治理的时代诉求 [J/OL]. 电化教育研究，2022（06 vo 43）：32-38. DOI：10.13811/j.cnki.eer.2022.06.005.

[16]SOUTHGATE E. Artificial intelligence, ethics, equity and higher education[R]. Technical Report. National Centre for Student Equity in Higher Education …, 2020.

[17]ZHOU X, VAN BRUMMELEN J, LIN P. Designing AI learning experiences for K-12: emerging works, future opportunities and a design framework[J]. arXiv preprint arXiv:2009.10228, 2020.

[18]JOSHI S, RAMBOLA R K, CHURI P. Evaluating artificial intelligence in education for next generation[C]// Journal of Physics: Conference Series: 卷 1714. IOP Publishing, 2021: 012039.

[19]POPKHADZE N. The Good, The Bad and The Ugly: AI in the higher education[J]. Hallinnon Tutkimus, 2021, 40(4): 254-263.

[20] 侯浩翔. 人工智能时代学生数据隐私保护的动因与策略 [J]. 现代教育技术，2019，29（06）：12-18.

[21] 赵磊磊，张黎，代蕊华. 教育人工智能伦理：基本向度与风险消解 [J/OL]. 现代远距离教育，2021（05）：73-80. DOI：10.13927/j.cnki.yuan.20210831.005.

[22] 苟学珍. 智能时代教育数据安全的伦理规约简论 [J/OL]. 电化教育研究，2021，42（09）：35-40. DOI：10.13811/j.cnki.eer.2021.09.005.

[23]CAMPOLO A, SANFILIPPO M R, WHITTAKER M, 等. AI now 2017 report[J]. 2017.

[24]COWIE R. Ethical issues in affective computing[J]. The Oxford handbook of affective computing, 2015: 334-348.

[25]XIAO Y, WATSON M. Guidance on conducting a systematic literature review[J]. Journal of Planning Education and Research, 2019, 39(1): 93-112.

[26] 胡航，杨旸. 多模态数据分析视阈下深度学习评价路径与策略 [J/OL]. 中国远程教育，2022（02）：13-19+76. DOI：10.13541/j.cnki.chinade.2022.02.007.

[27] 卢宇，汤筱玮，宋佳宸，等. 智能时代的中小学人工智能教育：总体定位与核心内容领域 [J]. 中国远程教育，2021（5）：11.

[28] 刘凤娟，赵蔚，姜强，王磊. 基于知识图谱的个性化学习模型与支持机制研究 [J]. 中国电化教育，2022（05）：75-81+90.

[29] 田贤鹏. 隐私保护与开放共享：人工智能时代的教育数据治理变革 [J/OL]. 电化教育研究，2020，41（05）：33-38. DOI：10.13811/j.cnki.eer.2020.05.005.

[30] 冯锐，孙佳晶，孙发勤. 人工智能在教育应用中的伦理风险与理性抉择 [J/OL]. 远程教育杂志，2020，38（03）：47-54. DOI：10.15881/j.cnki.cn33-1304/g4.2020.03.005.

[31]ZHOU Y. 舍恩伯格大数据教育应用思想的伦理关怀 [J]. 中国电化教育，2020（12）：55-62.

[32] 李晓岩，张家年，王丹. 人工智能教育应用伦理研究论纲 [J/OL]. 开放教育研究，2021，27（03）：29-36. DOI：10.13966/j.cnki.kfjyyj.2021.03.003.

[33] 刘三女牙，杨宗凯，李卿. 教育数据伦理：大数据时代教育的新挑战 [J]. 教育研究，2017，38（04）：15-20.

[34] 高山冰，杨丹. 人工智能教育应用的伦理风险及其应对研究 [J]. 高教探索，2022（01）：45-50.

[35] 吴河江，涂艳国，谭轼纱. 人工智能时代的教育风险及其规避 [J]. 现代教育技术，2020，30（04）：18-24.

[36]DRACHSLER H, GRELLER W. Privacy and analytics: it’s a DELICATE issue a checklist for trusted learning analytics[C]//Proceedings of the sixth international conference on learning analytics & knowledge. 2016: 89-98.

[37] 王萍，田小勇，孙侨羽. 可解释教育人工智能研究：系统框架、应用价值与案例分析 [J/OL]. 远程教育杂 志，2021，39（06）：20-29. DOI：10.15881/j.cnki.cn33-1304/g4.2021.06.003.

[38] 刘三女牙，刘盛英杰，孙建文，等.智能教育发展中的若干关键问题[J/OL].中国远程教育，2021（04）：1-7+76. DOI：10.13541/j.cnki.chinade.2021.04.001.

[39]HOGENHOUT L. A framework for ethical AI at the United Nations[J]. arXiv preprint arXiv:2104.12547, 2021.

[40] 王萍，田小勇，孙侨羽.可解释教育人工智能研究：系统框架、应用价值与案例分析[J/OL].远程教育杂　志，2021，39（06）：20-29. DOI：10.15881/j.cnki.cn33-1304/g4.2021.06.003.

[41]PARRISH J L. PAPA knows best: Principles for the ethical sharing of information on social networking sites[J]. Ethics and Information Technology, 2010, 12(2): 187-193.

[42] 杜静，黄荣怀，李政璇，等.智能教育时代下人工智能伦理的内涵与建构原则[J/OL].电化教育研究，2019，40（07）：21-29. DOI：10.13811/j.cnki.eer.2019.07.003.

[43] 赵磊磊，张黎，代蕊华.教育人工智能伦理：基本向度与风险消解[J/OL].现代远距离教育，2021（05）：73-80. DOI：10.13927/j.cnki.yuan.20210831.005.

[44] 冯锐，孙佳晶，孙发勤.人工智能在教育应用中的伦理风险与理性抉择[J/OL].远程教育杂志，2020，38（03）：47-54. DOI：10.15881/j.cnki.cn33-1304/g4.2020.03.005.

[45] 邓国民，李梅.教育人工智能伦理问题与伦理原则探讨[J/OL].电化教育研究，2020，41（06）：39-45. DOI：10.13811/j.cnki.eer.2020.06.006.

[46] 刘三女牙，刘盛英杰，孙建文，等.智能教育发展中的若干关键问题[J/OL].中国远程教育，2021（04）：1-7+76. DOI：10.13541/j.cnki.chinade.2021.04.001.

[47] 赵慧臣，唐优镇，马佳雯，等.人工智能时代学习方式变革的机遇、挑战与对策[J].现代教育技术，2018，28（10）：20-26.

[48]ZHOU Y.舍恩伯格大数据教育应用思想的伦理关怀[J].中国电化教育，2020（12）：55-62.

[49] 庞茗月，胡凡刚.从赋能教育向尊崇成长转变：教育大数据的伦理省思[J/OL].电化教育研究，2019，40（07）：30-36+45. DOI：10.13811/j.cnki.eer.2019.07.004.

[50] 安涛.“算计”与“解蔽”：人工智能教育应用的本质与价值批判[J].现代远程教育研究，2020，32（06）：9-15.

[51] 胡小勇，黄婕，林梓柔，黄漫婷.教育人工智能伦理：内涵框架、认知现状与风险规避[J].现代远程教育研究，2022（02 vo 34）：21-28+36.

[52]CELLAN-JONES R. Stephen Hawking warns artificial intelligence could end mankind[J]. BBC news, 2014, 2(10): 2014.