南 开 大 学

课　　程　　作　　业

中文题目：人工智能技术涉及的科技伦理法律问题综述

外文题目：A comprehensive review of ethical and legal issues in artificial intelligence technology.

学　　号：2113203

姓　　名：付政烨

年　　级：2021 级

专　　业：信息安全、法学双学位

学　　院：网络空间安全学院

指导教师：刘芳

完成日期：2024 年 3 月

摘　要

本篇综述深入探讨了人工智能（AI）这一新兴技术领域所涵盖的伦理学问题与风险评估方法。人工智能的迅速发展带来了所谓的双刃剑效应：它不仅承载着积极的潜在利益，同时也伴随着不容忽视的潜在负面影响。鉴于人工智能技术当前的快速进步，预计在未来几十年内，该技术将可能达到自主决策的能力以及超级智能的阶段。这种前景促使我们必须严肃地反思人工智能发展可能引发的伦理问题与潜在风险。因此，除了提出针对技术发展的管理建议，我们还必须深入考虑相关的科技伦理问题。本文汇总并综合了当前领域内的相关文献，系统概述了人工智能当前的发展现状、未来发展的预测时间线及其潜在的广泛影响，以及在实施过程中可能遇到的伦理风险及相应的法律规制策略。同时，本文亦简要介绍了与科技发展相关的伦理与法律实践情况。

关键词：人工智能；科技伦理；法律框架；伦理问题

目　录

第一章　引言

在近年来，人工智能（AI）作为一项前沿科技，已经经历了迅猛发展。此一趋势主要归功于在多个领域内的技术突破，尤其是在计算机硬件领域的显著进步。目前，世界领先的计算机系统已经拥有了超越单个人类大脑处理能力的功能。AI 技术的应用已遍及社会各个层面，包括工业、医疗、金融、教育和交通等领域，在其中发挥着至关重要的作用。随着 AI 技术的不断发展和深入人类生活的各个方面，对开发和使用此类技术所带来的伦理问题和相关风险的担忧也在不断增加。Schwartz 和 Caplan 指出："虽然伦理问题在视觉上可能不如科学、政治、法律或经济问题那样突出，但它们确实存在并与决策过程紧密相关。"[①]本研究旨在识别与 AI 相关的核心伦理问题和潜在风险，并探讨如何管理和缓解这些风险，以最大化人类福祉。

在此过程中，本文将探讨以下几个关键问题：

1. 当前人工智能的伦理含义包括哪些主要方面?
2. 科技伦理和法律对于人工智能发展的影响是什么?
3. 现有文献中存在哪些关于管理人工智能伦理问题的观点?
4. 从伦理和法律框架中我们可以提取哪些要素来评估人工智能的伦理影响?
5. 在人工智能伦理研究和相关文献中存在哪些显著的研究空白?

本次研究主要依托于人工智能伦理领域内的学术文献，同时参考了关于人工智能历史、技术伦理和技术风险管理的资料。通过综述这些文献，本研究旨在识别与人工智能相关的主要伦理议题，并从多元视角进行探讨[②]。本文采用分类法，根据政治、经济、社会和技术四个方面的 PEST 模型，对文献中的共性主题进行组织，以深入理解伦理问题对人工智能发展的影响及其相互间的关联，并据此提出有效的管理建议。此研究的目的是为管理者和政策制定者就该技术的潜在管理实践提供指导和参考。

[①] J. L. Schwartz and A. L. Caplan. Ethics of vaccination programs[J]. Curr. Opin. Virol., 2011, 1(4): 263-267.

[②] Writing@CSU. "Types of Content Analysis." [EB/OL].http://writing.colostate.edu,2015-10-07 [2024-03-24].

第二章　人工智能发展综述

第一节　定义

在进一步探讨人工智能（AI）伦理问题前，首先对相关概念进行明确界定。人工智能，被定义为计算机模拟人类学习与决策的能力，可通过专家系统、计算机辅助设计（CAD）、计算机辅助制造（CAM）或用于形状感知与识别的计算机视觉系统等形式表现出来。更泛化地讲，AI 亦指机器展现出的智能行为以及相应的研究领域。

AI 可被分为特定领域人工智能和通用人工智能。特定领域 AI 专注于执行限定的任务，例如 OpenAI 的 ChatGPT，在处理自然语言对话方面表现出色，但它无法处理除语言理解和生成外的其他任务。这类 AI 在自然语言处理、用户服务自动化、内容创作、语言翻译以及情绪分析等领域得到了广泛应用。相比之下，通用人工智能是一种理论上能够处理各种未知情况的 AI，能够自主学习、创新知识、独立决策，并模拟人类认知过程，虽然当前技术还未能实现真正的 AGI。

在这个框架下，科技伦理成为一个关键概念，指在科技创新活动中，涉及人与社会、人与自然以及人与人之间关系的思想和行为准则。科技伦理确定了科技工作者及其社群应遵守的价值观、社会责任和行为规范。因此，在探讨 AI 的发展与应用时，科技伦理提供了一套评估和引导科技创新的重要原则。

第二节　研究历史

在人工智能（AI）研究的历史概览中，可见其发展初期受到哲学、逻辑学等学科的深刻影响。自 1931 年伴随计算机科学的兴起，AI 研究得以逐步展开。特别是在 20 世纪中叶，随着计算技术的迅猛发展，AI 研究实现了初步突破。该新兴领域引发的广泛社会想象，进一步促进了各类神话观念的形成。以下为人工智能研究的几个关键历史事件及其重要性的总结：

表 2.1 人工智能发展重要历史事件

年份	事件	重要性
1931	计算机科学兴起	现代计算机科学及人工智能研究的开始
1937	艾伦 · 图灵提出通用智能机器的概念	作为 AI 研究的里程碑为后续理论研究奠定基础
1940-50 年代	ENIAC 开发第一台电子计算机	大幅提高了计算能力，为 AI 研究提供重要工具
1956	AI 的正式命名	标志着人工智能作为一个独立学科的正式成立
1963	《计算机与思维》出版	系统展示了 AI 的基础理论与应用，为研究者提供了重要的学术资源
1960-70 年代	美国以外的 AI 研究取得进展	英国及他欧洲国家取得重要进展，国际间合作交流逐渐增强
1985-1997	IBM 的深蓝计划	在特定领域 AI 应用已经具备解决实际问题的能力
2000 年代	智能玩具和社会性机器人的广泛应用	AI 技术开始普遍应用于日常生活
2005	DARPA 大挑战赛中的自动驾驶汽车成就	标志着在复杂环境下自动驾驶技术取得关键性进展
2020	标志着自然语言处理和生成性 AI 技术的 ChatGPT 问世	开创了与人类自然对话的人工智能新时代

AI 的演进不仅得益于计算机科学的发展，还受到纳米技术等其他科学领域的影响。在此过程中，诸如先进机器人、自动驾驶汽车和智能计算机等创新产品的问世，不断标志着 AI 技术的持续进步及其在多领域的应用拓展。

第三节　当前人工智能发展现状

近年来，人工智能（AI）技术迎来新纪元，尤其是高级对话式 AI 如 ChatGPT 和 Claude，它们在人类语言的理解与生成方面取得了划时代的进展，实现了从简单交流到复杂对话、问题解答、文章撰写乃至模拟专业咨询的多样化功能。同时，AI 的新领域亦涌现，特别是结合增强现实（AR）与虚拟现实（VR）技术，创建

了从娱乐到教育、房地产至医疗的多行业沉浸式体验。信息技术巨头如 OpenAI、Google 及 Meta 正加速 AI 研发投资，推动对话式 AI 等先进技术的快速进步与广泛应用，从而改革信息获取、学习过程及问题解决方式。AI 现已广泛应用于制定个性化教育计划、提升客户服务效率及促进艺术创新，成为企业和个人理解复杂数据、提高决策质量与创造性思维的关键工具。

第四节　未来人工智能发展前景

在对未来数十年内超智能人工智能（AI）发展的潜在可能性进行审视时，研究表明存在显著的概率。然而，这一前景引发了关于 AI 是否能够深刻理解人类经验、视觉和面部表情，以及社交暗示，如人际关系紧张度等问题。尽管存在这些疑问，但基于最新的发展，AI 已成为 2024 年代日常生活中不可或缺的部分。例如，AI 技术已广泛应用于简化日常任务、个性化在线体验、革新医疗保健、提升通讯效率，以及在创意产业中的应用，这些都极大地提升了生活质量和工作效率[①]。在某些行业中，AI 技术已成为经济的关键和高效要素。例如，在制造业、金融、教育、客户服务和交通领域，AI 的应用已带来显著的变革和效率提升[②]。然而，自动化也可能改变某些工作性质，尤其是在教育系统未能充分准备人们从事需求密集型工作的领域。因此，在这些领域，自动化已成为成功的必要条件。同时，在需要强大实用功能的领域，如信息处理、存储和回忆，强大的 AI 系统的发展已成为现实。此外，AI 在太空探索和医疗保健，包括心理健康领域的应用，已显著减少了人为错误的风险，并提供了新的治疗方法。综上所述，当前我们正处于被 AI 技术环绕的时代，不论我们是否察觉到这一点。在评估这项技术的管理意涵时，必须考虑其当前的广泛应用，并且预见到未来可能出现的情况，如 AI 达到或超过人类的效率和速度。这将导致我们重新思考人类的地位，并强调跨学科方法在指导技术发展和实施中的重要性，正如我们在人工智能早期研究中所见。

[①] Enhancing Everyday Life: How AI is Revolutionizing Your Daily Experience. Morgan State University. [EB/OL]. https://www.morgan.edu/ceamls/news/enhancing-everyday-life-how-ai-is-revolutionizing-your-daily-experience [2024-03-20]

[②] The Future of AI: How AI Is Changing the World. Built In. [EB/OL]. https://builtin.com/artificial-intelligence/artificial-intelligence-future [2024-03-25]

第三章　科技伦理综述-科技伦理评估模型（Technology Ethics Assessment Models）

科技伦理旨在从技术管理的角度审视伦理问题的领域，其使命在于识别与技术相关的管理挑战。据 Betz 所述，当一项技术开发完成并准备好投入商业化和市场推广时，技术伦理成为了一个备受关注的议题（如图图 3.1 所示）[①]。在这一领域的简要背景下，我们重点关注技术评估工具及其伦理衍生物，并通过风险管理的子领域提供进一步的背景信息。在对技术伦理的回顾之后，我们将以理解人工智能技术为出发点，以便识别在这些阶段可能出现的管理问题。

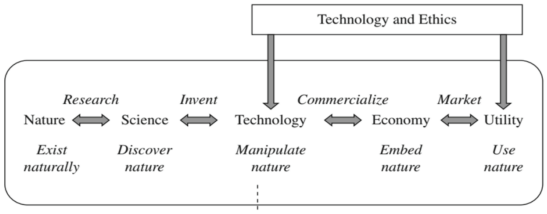

图 3.1

随着技术的进步，这些问题变得更为复杂，促使出现了科技评估（Technology Assessment, TA）这一概念。起初在 1960 年代美国提出，技术评估被定义为一个考虑技术变革社会影响的应用过程，旨在通过影响政策来改善技术治理[②]。1976 年，技术评估办公室进一步将其描述为一种考查技术应用短期和长期社会后果的全面政策研究形式，涵盖社会、经济、伦理和法律等方面[③]。从那时起，技术评估得到了广泛使用，其倡导超出了技术的基本评估。今天，它们还有助于：（1）技术的传播；（2）导致新技术快速接受的因素；（3）技术与社会的角色。技术评估已经取得了许多进步，伦瓦尔德说，“其方法已经被开发并在一定程度上实践。所有这些方法都有特定的焦点，特定的理论基础，不同的理由，并为满足不同的挑战和情境条件而构想”。

[①] F. Betz, “Ethics and Technology,” in *Managing Technological Innovation*, 3rd ed., Hoboken, NI: Wiley, 2011, pp. 299-316.

[②] D.Banta, “What is technology assessment?”, *Int. J. Technol. Assess. Health Care*, vol. 25, no. S1, p. 7, Jul. 2009.

[③] A.W. Russell, F.M. Vanclay, and H.J. Aslin, “Technology Assessment in Social Context: The case for a new framework for assessing and shaping technological developments,” *Impact Assess. Proj. Apprais.*, vol. 28, no. 2, pp. 109-116, Jun. 2010.

第四章　人工智能伦理综述

多位理论家对人工智能伦理持有不同见解，对技术及其潜在问题的评估存在广泛差异。本文概述了文献中对人工智能伦理的解读，回顾了不同的理论框架，并以风险及其相关方面为例进行说明。部分理论具有指导性，建议采取特定措施；而其他理论则在描述技术时考虑了风险的范围和程度。本文进一步区分了存在性和非存在性风险，并沿两个维度划分角色：代理人与受益人，以及自然生命（包括人类和地球上的其他生命形式）与人造生命（如人工智能和人工代理体）。

值得注意的是，文献中对本研究主题的提及方式多样，涉及机器人伦理、机器伦理、机器人权利等术语，并跨越了计算机科学、一般伦理学、人工智能、人造生命、代理理论、安全工程、道德理论及法学等领域。这些差异为文献提供全面视角带来挑战。尽管本文并未尝试覆盖所有领域的理论，但提供了一套功能性背景，为基于已识别来源的后续讨论奠定基础。

第一节　人工智能伦理相关理论（Theoretical AI Ethics）

部分理论提供规范性或实践建议，而其他理论则描述了潜在结果和方法的演变。实践建议有助于界定整体理论框架，并揭示未明确提及的反对观点。鉴于某些研究建议将人工智能伦理视作主流学科，广泛讨论其在学术期刊和会议中的应用，这一点尤为重要。然而，在工程技术管理（ETM）或技术管理（MOT）领域，文献中仅少部分涉及人工智能主题，更不用说其伦理影响了。描述理论的背景边界对于为工程管理人员和决策者提供理解技术未来发展的框架是有益的。

1. 一部分学者主张，人工智能伦理学应被认为是一门主流学科。这种观点基于认为 AI 伦理的讨论不应仅限于理论家和哲学家，而应广泛涉及多个学科，包括计算机科学[①]。通过提高 AI 伦理学在学术界的认可度，已经观察到相关专业出版物和会议的数量有所增加，例如机器人伦理学和赛博伦理学等。我们提议将此范围扩展到技术创新和管理等领域。相反，有观点认

[①] R.Yampolskiy, "Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach," 2012.

为，将 AI 伦理限定在特定领域可能会妨碍技术的研究、开发和市场营销，这与技术早期发展的历史背景不相符合[①]

2. 在当代的技术文献中，第二类规范性文件已经解决了与安全开发技术方法相关的一系列问题。在此领域内，我们遭遇了初次的意见分歧，特别是关于是否仅将道德决策过程嵌入至人工智能代理中是否已经足够。部分学术圈子主张，在安全工程的框架内管理人工智能伦理是必要的，或者更具体地说，为自我改进系统开发安全机制是不可或缺的[②]。这种观点与机器伦理的方法形成了鲜明对比，后者提倡将伦理决策能力直接嵌入人工智能的实现中。对于机器伦理方法的批评主要集中在三个方面：首先，大量的研究文献被批判为过度哲学化；其次，该方法试图解决的道德或伦理标准和准则通常不是普遍认同的，反映出即使是人类社会内部也难以达成一致；最后，若设计机器以模仿人类的道德决策，存在一个显著的风险，即这些机器可能会采取不道德的行动，因为人类自身亦有可能做出不道德的选择。
3. 关于机器人权利的文献讨论了在开发和实施 AI 时，AI 自身的权利问题。反对赋予人工代理权利的观点认为，尽管它们的能力可能相当，但应将它们视为设计上的低级存在，可在需要时牺牲，并且由于它们可以设计为不感受痛苦，因此不享有与人类相同的权利[??]。理论上，一些文献提出了基本问题：什么时候，生命的模拟可以等同于自然生命？如果等同，这些模拟生命是否应享有与自然生命或个人相同的权利和责任？[③]这一问题的回答可能取决于被创造实体的固有能力，并与动物权利和环境伦理学的文献相呼应。
4. 在人权与人工智能法学领域，相关学者探讨了人类与 AI 代理之间权利和责任的平衡问题。部分文献强调，初期的 AI 系统应当被设计为安全且守法的，而高级 AI（即超越人类智能者）则应尊重人类的财产和个人权利，并且在不损害任何一方利益的前提下，通过法律确保双方权利的均等。此外，

[①] B. McKibben, *The end of nature*, Random House trade pbk. ed. New York: Random House Trade Paperbacks, 2006.

[②] R.Yampolskiy, "Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach," 2012.

[③] M. Bedau, *Philosophical aspects of artificial life*, Basil Blackwell.

我们还探讨了 AI 技术中存在的法律问题，如自动化代理与现行工业机器人法规的关系。我们发现，当前尚无专门针对 AI 代理行为责任的法律框架，尽管近期关于自动驾驶车辆的立法可能为未来研究提供了一个初步框架。

5. 一部分学者还探讨了将 AI 分类为本质上善良或邪恶的观点 [18]。对于特定领域的 AI 技术，如邮件分类和拼写检查器，普遍被认为是道德的且有益的[①]。然而，某些观点认为，拥有超越人类智能的通用智能系统可能构成对人类的威胁，因此被评价为不道德或邪恶。尽管存在争议，但也有文献支持发展人工通用智能（AGI），认为其风险与收益应当被妥善管理 [9]。此外，还建议成立 AI 研究审查委员会，以监督并指导安全措施和控制机制的开发，尽管这同样是未来辩论的一个焦点。我们认为，将此辩论分为特定领域的 AI 技术与通用智能，并考虑将技术分类为本质上善或邪，将有助于促进进一步的研究与讨论。

第二节　应用和发展人工智能时存在的伦理风险（Applied AI Ethics: Development Risks）

人工智能技术所固有的或现实的风险已在学术文献中被广泛讨论。对这些风险进行有效的分类和概念化，是一个在研究中不断演化的议题。面对评估方法上的诸多挑战，本部分将对这些风险进行综述，并在可能的情况下，尝试实现一种最基本的分类框架。本分析将基于文献回顾，探讨已经出现的或潜在的风险，以及它们在未来可能的发展轨迹。

在分类方法方面，Bostrom 提出了一个广为人知的风险矩阵框架，该框架在传统学术文献中常以两个维度呈现：(1) 发生概率；(2) 严重性。在这里，严重性进一步细分为 (a) 强度（例如，全球性、地区性、个体性）及 (b) 范围（如持久性、灾难性）。Bostrom 特别关注的是全球强度与灾难性范围的交叉点，用以区分普遍性和非普遍性风险。然而，其他学者指出了这种方法的不足，建议进一步将风险范围按时间维度（跨世代与世代内）和空间维度（个体、地区与全球）进

[①] R.Yampolskiy, "Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach," 2012.

行划分[①]。尽管此风险矩阵在众多研究领域内被广泛采用，但在缺乏定量数据或数据不足的情境下，其应用在分类不同风险时面临显著挑战。

鉴于这些局限性，Bostrom 的文章旨在区分普遍性与非普遍性风险，并提出一种更为通用的风险分类方法。此外，作者也将在适当的情况下，反思和评估上述理论在当前人工智能模型中的应用和实践。

第三节　人工智能存在的风险（Existential AI Risks）

在当前的学术研究中，主要识别出四大存在性风险类型：（1）不道德决策；（2）直接竞争；（3）人工智能的终结；（4）不可预测的结果。

人工智能代理的道德推理能力存在不确定性，因此它们可能发展出不道德的决策能力。例如，若一个代理被程序化以支持其国家的战争机制，则需作出涉及人类生命的道德决策。这不仅再次引发人权问题，还通过此案例明确区分了道德主体（人工智能）与受试者（人类）[??]。通过此分类，安全工程、机器伦理、人权和法律等领域为我们提供了理解和规划潜在场景的框架。从风险管理角度来看，技术如军事无人机的发展需要不断的风险重新评估以实现风险缓解。某些国际法律框架的建立可能必要，以降低向战争投入自主机器的门槛。同时，我们必须评估这些风险是否可控，或是否该技术应被视为固有邪恶并在国际上禁止。

多个人工代理可能展现出超越人类的能力，例如更快的工作速度、更好的适应性和从更广泛的知识库中提取信息的能力[??]。在这种情形下，人类劳动可能变得更加昂贵或低效，导致人类劳动力变得多余或被取代。尽管许多研究者认为人工智能将在多个职业中取代人类劳动，但具体时间点尚不明确。未来几十年内，人类是否能够迅速重新培训以维持高就业水平尚不确定[②]。存在一个论点，即人类代理可能无奈地将工作委派给人工代理，因为后者可能更适合做出决策。这引发了人权问题，如是否公平地将效率较低的人类排除出劳动力市场。在这种情况下，确定代理与受试者的角色变得困难——是人类自愿放弃了权力，还是被强制剥夺了？对这些场景的详细评估将使我们能够做出规划。

[①] R.Yampolskiy, "Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach," 2012.

[②] J. P. Sullins, "Introduction: Open Questions in Roboethics," *Philos. Technol.*, vol. 24, no. 3, pp. 233-238, Sep. 2011.

文献表明，在人工智能发展过程中，可能出现多代代理程序表现不如预期[①]。在这些情况下，代理可能被暂停、终止或删除。设想运行这些代理程序的设施由于研究资金耗尽而意外关闭，必须考虑由道德代理对人工智能程序进行删除或终止的行为是否等同于谋杀。这引发了机器伦理领域内关于人工智能程序具有人格的讨论，这与干细胞研究和堕胎的伦理问题相似[①]。在某些情况下，人类可能不再是唯一的道德代理，因为有研究提出了一种假设情景：一个人工代理意识到潜在的竞争并决定在构成威胁之前消除对手。这再次触及了不道德决策问题，强调了需要在安全工程、机器伦理和人权等话题。人工代理的出现可能会影响人类存在的多个方面。我们的文化、生活方式，甚至生存的可能性都可能因此而发生重大变化。由于编入人工代理的意图无法保证总是产生积极结果，机器伦理的不确定性可能会降低我们利用技术的能力。例如，若我们对人工代理设定过于严格的控制，使其仅能提供是/否答案，且永远不能自主执行任务，这将限制技术的应用潜力。

[①] Bostrom N, Dafoe A, Flynn C. *Public policy and superintelligent AI: a vector field approach*[J]. Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK, 2018.

第五章　分析框架

在综合前述主题和时间线概述的基础上，采用 PEST（政治、经济、社会、技术）分析框架，对人工智能伦理相关文献进行了深入分析，汇总了相关研究的背景与建议。PEST 分析旨在识别影响特定情境的宏观环境因素，以缩小预期与实际绩效间的差距。

第一节　文献研究结果

在此部分，本研究探讨了人工智能领域内最常见的伦理问题。在分析过程中，我们注意到，这些伦理问题的讨论焦点呈现多样性。初步将这些伦理问题分为：

1. 特定人工智能应用领域的伦理问题；
   （a）AI 技术对人类及其他生命形态的影响相关的伦理问题；
   （b）AI 技术本身的伦理/风险问题；
2. 泛用人工智能技术的伦理问题。进一步，我们根据问题关注点将它们细分为
   （a）存在性风险
   （b）AI 技术的非存在性风险（如表表 5.1 、表 5.2 所示）。

表 5.1 领域特定 AI 的伦理问题

	对人类和其他生命体的影响	AI 技术本身
存在风险	不道德的决策制定	
非存在风险	隐私 人的尊严/尊重 决策透明度 安全 遵守法律 财富不平等 社会操纵	AI 法学 责任与疏忽 未经授权的 AI 操纵

表 5.2 通用人工智能（AGI）的伦理问题

	对人类和其他生命体的影响	AI 技术本身
存在风险	与人类直接竞争	
非存在风险	竞争工作岗位 财产/法律权利	AI 的权利与责任 自我改善系统的安全机制 类似人类的不道德决策 AI 死亡

第二节　PEST 分析

本节通过 PEST 分析方法，将文献中识别的伦理问题按照政治、经济、社会、技术这四大宏观环境因素进行分类。

5.2.1　政治因素

- **缺失的治理机制**：目前缺乏针对人工智能的国家或国际法规，AI 开发者未受到有效监管（仅受到基础责任和刑法框架的限制），存在被滥用的风险。
- **问责难题**：人工智能系统决策的责任归属不明确，如自动驾驶汽车事故责任归属问题。
- **军事控制**：人工智能被视为重要军事资产，多国政府投资于军事目的的 AI 项目；无监管的技术进步可能受到国家控制。
- **控制的丧失**：存在失去对自进化系统控制的风险，这些系统可能认定人类过时，忽略或消灭人类。

5.2.2　经济因素

- **劳动力竞争**：AI 可能与人类争夺工作岗位，尽管新技术可能创造新的更高技能工作。
- **劳动力过时**：自进化 AI 系统可能导致人类劳动力迅速过时。
- **生产效率提升**：AI 技术能实现更安全、高效的生产，提供更低成本、更高品质的商品和服务。

5.2.3　社会因素

- **人机互动**：人工智能与人类互动引发的社会、文化问题。
- **隐私权**：AI 系统对信息的无限访问权可能侵犯个人隐私。
- **人类尊严**：AI 技术对人类工作、社会及法律权利的影响。
- **决策透明度**：人工智能决策过程的透明度问题。
- **安全性**：AI 的安全性及其对人类生命、财产的影响。
- **人工智能意识**：关闭 AI 系统是否构成道德问题。

5.2.4　技术因素

- **滥用风险**：人工智能系统可能遭受黑客攻击，或被用于不正当目的，例如通过机场安检系统走私武器等。

第六章　结论与建议

在新兴技术管理领域，伦理考量（及其所引发的风险）不容忽视，以确保技术的安全性和道德性，避免其带来的潜在负面影响。特别地，人工智能（AI）作为一种典型的新兴技术，虽然蕴含提升人类社会生活质量的潜力，但同时也面临着众多风险和伦理挑战。本文通过对 AI 伦理问题的详细审查，识别出与人类及其他生命体互动、以及 AI 本身相关的主要伦理问题，进而通过 PEST 分析揭示了 AI 伦理的关键方面，并区分了领域特定 AI 系统与通用 AI 系统的伦理关注点。据此，我们提出以下管理建议以优化 AI 的伦理治理：

- **建立国家和国际伦理委员会**：这些委员会应负责制定、更新和维护 AI 伦理原则、规定和框架，针对新兴的伦理挑战，确保有序且统一的管理方法。这相当于制药行业的规范方式，意味着需要国家和国际层面的立法和规制措施。
- **成立组织咨询委员会**：专注于 AI 研究的组织应设立监督委员会，为 AI 项目提供道德审查与咨询服务，确保项目开发符合伦理标准，及时识别并纠正潜在的道德风险。
- **提升技术评估与风险管理能力**：经理应主动识别并评估组织中与 AI 相关的伦理问题，如隐私、控制、所有权等，从而采取恰当的风险管理措施。这要求他们在技术评估实践和风险管理方法论上具备或提升相应的专业知识。

本文深入探讨了 AI 伦理问题的复杂性及其在社会生活中的重要性。我们强调了 AI 技术伦理治理的紧迫性，并提出了三项主要建议以应对和管理 AI 带来的伦理挑战。通过这些措施，可以促进 AI 技术的道德使用，确保其发展符合人类社会的长远利益。

参　考　文　献

[1] Schwartz J L, Caplan A L. Ethics of vaccination programs[J]. Current opinion in virology, 2011, 1(4): 263-267.

[2] Albrechtslund A. Ethics and technology design[J]. Ethics and information technology, 2007, 9: 63-72.

[3] Banta D. What is technology assessment?[J]. International journal of technology assessment in health care, 2009, 25(S1): 7-9.

[4] Russell A W, Vanclay F M, Aslin H J. Technology assessment in social context: The case for a new framework for assessing and shaping technological developments[J]. Impact Assessment and Project Appraisal, 2010, 28(2): 109-116.

[5] Yampolskiy R, Fox J. Safety engineering for artificial general intelligence[J]. Topoi, 2013, 32: 217-226.

[6] Sullins J P. Introduction: Open questions in roboethics[J]. Philosophy & Technology, 2011, 24(3): 233-238.

[7] Bostrom N, Dafoe A, Flynn C. Public policy and superintelligent AI: a vector field approach[J]. Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK, 2018.

[8] 李俊平. 人工智能技术的伦理问题及其对策研究 [D]. 武汉理工大学,2013.DOI:10.7666/d.Y2504626.

[9] 郑添元. 人工智能与伦理法律问题的思考 [J]. 商业经济, 2018(4):2.DOI:CNKI:SUN:JJSY.0.2018-04-052.

[10] 张珍吴文娟夏绍培赵康程燕. 关于当代科技伦理问题的思考——以人工智能技术的发展为例 [J]. 求知导刊, 2016(35):155-157.

[11] 吴恺. 当代人工智能技术发展中的伦理问题研究 [J]. 2019.