

数智健康国际动态

北京市卫生健康大数据与政策研究中心

2025. 10. 30

(十) 数智伦理

随着人工智能技术的迅猛发展，其在医疗领域的应用正以前所未有的深度和广度重新塑造现代医学的面貌。从精准的影像诊断到个性化的治疗方案，从高效的管理流程到远程的医疗服务，人工智能为提升医疗质量与效率带来了革命性的机遇，但与此同时，也使医学伦理面临着严峻挑战。面对机遇与挑战，我们并非束手无策，可以通过健全的法规、向善的技术和坚守的伦理原则，将技术创新牢牢锚定在人文关怀的基石之上，共同引导人工智能医疗迈向一个更加公正、可信且充满温度的明天。这不仅是技术发展的必然要求，更是我们对生命尊严与健康权利的基本承诺。以下两篇文章将从理论与实践两个视角探讨如何通过规范、技术与人文关怀的平衡构建可信、公正且高效的智慧医疗环境。

第一篇文章主要从理论角度系统探讨了人工智能（AI）在医疗中的应用、潜在风险及伦理挑战。文章指出 AI 技术通过深度学习和大数据分析，在辅助诊断、预测医学、临床决策支持、远程医疗及手术机器人等领域展现出巨大潜力。例如，AI 在放射学和皮肤病学中能够快速识别病变，提高诊断准确性和效率；在预测医学中，AI 有助于个性化预防和优化治疗路径。此外，生成式 AI 可协助起草医疗报告，减轻医生行政负担，使其更专注于患者沟通。然而，其应用也引发了一系列伦理问题，比如医患信任与透明度，AI 决策过程的“黑箱”特性可能导致患者对诊断结果的不信任，而医生有责任向患者解释 AI 工具的原理、局限性和使用目的，以维护知情同意和医患信任；再比如数据隐私与保密性，AI 系统依赖大量患者数据进行训练和验证，存在数据泄露和滥用的风险，而保护患者隐私不仅是法律义务，也是医学伦理的核心；还有，医生去技能化与自主权，过度依赖 AI 可能导致医生诊断能力和临床判断力下降，甚至削弱其在决策中的主导地位；公平性与可及性，AI 算法若基于有偏数据训练，可能加剧对少数群体的不公；数字鸿沟可能使部分患者无法享受 AI 带来的医疗进步。文章还比较了不同国家对 AI 在医疗中的监管存在的显著差异：欧盟通过

《通用数据保护条例》和《人工智能法案》强调透明度、可解释性和患者权利，要求 AI 系统处于人类监督之下；美国缺乏统一的联邦法规，依赖行业指南和机构建议，强调 AI 的辅助角色而非替代医生；而中国则侧重于数据安全和技术规范，但在患者知情权和算法透明度方面的规定较为宽松。文章最后指出：总之，AI 既是医疗进步的推动力，也是伦理挑战的试金石。唯有在技术创新与人文关怀之间取得平衡，才能构建更加公正、高效且充满温度的医患关系。

第二篇文章主要通过对一个基于患者自主权的大型语言模型（LLM）案例的人机协议评估来具体阐述如何通过提示工程技术来提升患者与医生达成共识，以增强 LLM 在医学伦理中的可靠性与可信度。医学伦理有四大原则（仁慈、非恶意、患者自主、正义），其中患者自主权常被视为优先原则。随着大型语言模型在医疗教育、研究和实践中的广泛应用，其准确性、偏见与伦理适用性引发了广泛关注。文章先选取了 44 个涉及患者自主权的假设案例（包括同意能力、拒绝治疗、保密性、未成年患者同意等），由五名医生组成小组独立判断，并与三个基础 LLM（ChatGPT、LLaMA、Gemini）的回答进行比较。再采用多种提示工程技术（如思维链、N-shot 提示、问题细化、上下文扩展等）对 LLM 回答进行迭代优化，目标是提升其与医生共识的一致性。然后使用 Fleiss κ 和 Cohen κ 评估一致性与改进效果，用 McNemar 进行检验。其结果为：医生小组内部一致性高，基础 LLM 之间也存在基本一致性；基础 LLM 与医生共识的一致性较差；经过提示工程优化后，所有 LLM 与医生共识的一致性显著提升。最后得出结论：LLM 在医学伦理判断中具备潜力，但其初始表现与医生共识存在差距，提示其推理过程可能缺乏对伦理情境的深度理解。提示工程技术能有效提升 LLM 在专业领域的表现，尤其在结构化、情境化的任务中。人类专家的监督与干预仍是不可或缺的环节，尤其是在涉及患者价值观、情感与情境理解的复杂伦理决策中。本研究为 LLM 在医学伦理中的可控应用提供了实证支持，也为后续开发更具伦理敏感性的 AI 工具奠定了基础。但本项研究案例多集中在患者自主权，未覆盖全部医学伦理议题，也未使用检索增强生成或微调等进阶优化技术，未来可通过拓展案例范围、融合多模态数据，探索 LLM 在跨文化、多法律背景下的伦理适应性。

（徐健编辑）

译文一：

人工智能如何改变医患关系？医学伦理视角

Gianluca Montanari Vergallo, Laura Leondina Campanozzi, Matteo Gulino, Pasquale Ricci, Simona Zaami, Susanna Marinelli, Lorena Bassis, Vittoradolfo Tambone, Paola Frati, 徐健（译）

1. 简介

人工智能（AI）的核心概念自 1950 年以来就一直受到广泛争议，但直到 2000 年才由于深度学习算法的兴起和发展出现指数级增长，给该领域带来了重大突破。

虽然人工智能的定义仍然没有明确和普遍认可的定义，但有一个可行的定义：“人工智能系统”是指一种基于机器的系统，该系统旨在以不同程度的自主性运行，并且在部署后可能表现出适应性，并且对于显式或隐式目标，从其接收到的输入中推断出如何生成可以影响物理或虚拟环境的预测、内容、建议或决策等输出。这个定义让我们看到人工智能的巨大潜力，它能够基于如此大的数据存储来执行高度复杂的任务，以至于它们超出了任何人的分析能力。尽管这一前景潜力的实现涵盖了许多领域，但众所周知，医疗保健是此类应用广阔前景的领域之一。事实上，基于人工智能的新技术在这方面被证明特别令人鼓舞，例如，通过使用能够从经验和新信息中学习的经验机器学习（ML）算法，逐步提高其在诊断、治疗和预后阶段的表现，作为对医疗保健专业人员决策的支持。这构成了人工智能（AI）的基础。存在各种类型的机器学习，根据其数据分析方法和依赖程度进行分类。综上所述，机器学习（ML）分为三大类：识别模式的机器学习（无监督机器学习）、利用基于先前示例的算法进行分类和预测的机器学习（监督机器学习）以及利用奖励和惩罚系统制定旨在解决特定问题的解决方案策略的机器学习（强化学习）。人工智能在医学领域的最初重大成就始于分子医学中蛋白质复合物的预测，从而发现了新的药物靶点。电子病历（EMR）和医院数据库中包含的大量数据非常适合人工智能分析，提供有价值的见解。目前，电子病历很麻烦，缺乏有效的记录间通信。分析这个庞大的数据库非常有益。例如，人工智能可以识别有慢性病风险的个体，并加快和加强卫生系

统对成本效益比的计算，从而帮助决策过程。如今，人工智能被越来越频繁地用于分析和学习识别图像处理模式，引起了人们对放射学、病理学、眼科和皮肤病学等领域的极大兴趣——统称为“以视觉为导向的专业”。人工智能最大限度地减少常规临床实践中的常见错误并做出实时预测的能力非常重要和有效。

总体而言，人工智能有可能通过减少诊断和治疗的不确定性来增强医疗保健实践。它支持临床医生对患者病情进行预防、分类和分层；它预测康复是否需要特殊治疗，并通过数据收集和分析解释图像并检测人眼无法察觉的信号。此外，人工智能有助于了解疾病如何以及为什么发展；它有助于确定最合适的治疗方案，并通过提供更新的循证指南来帮助医生和医疗保健专业人员，即使是在病房内和外科手术过程中实时获取。

为了更好地了解这些引人入胜但有争议的技术的机会和风险，值得关注人工智能在支持诊断方面的应用，特别是在放射学和皮肤病学：已经设计了自动化系统，能够通过分析大量数据的算法来读取和识别图像，从而构成了快速、准确和精确地识别任何可疑病变的宝贵工具。良好的预测可以进行准确的诊断。事实上，人工智能是一种能够实施和开发那些能够及早发现疾病发作的“预测模型”的工具。在放射学的具体案例中，为了了解人工智能的使用和附加值，一项临时研究开发了一种能够分析和解释超过 10 万张胸部 X 光片的算法，最终证明在检测肺炎方面比专业评估和意见更准确。

人工智能应用也被用于各个医疗领域来辅助诊断。这些领域包括但不限于心脏病学、神经病学、肿瘤学、放射学、放疗、眼科、胃肠病学、妇科肿瘤学、乳腺病学、血液学和传染病。此外，此类模型在自闭症的诊断中显示出巨大的前景。由于自闭症的复杂性质，这尤其重要，它涉及不同的病因、严重程度、合并症以及与治疗管理相关的挑战。

鉴于人工智能系统似乎通过提高医疗保健质量预示着具体而广泛的好处，对这些做法在常规临床环境中的潜在缺点和后果的批判性评估不容忽视。应特别关注医学专业的基石：医患关系。

尽管目前尚不确定人工智能技术如何以及在多大程度上影响医患关系，但毫无疑问，无论是在护理体验还是临床结果方面，建立和维持良好的医疗保健关系对于提供有效的援助都至关重要。这种动态代表了现代以患者为中心的医学的基石，其

中医疗保健关系本身被认为是必不可少的，从而领先于和取代任何技术援助，并成为利用大量医学治疗能力的典型途径。

人工智能系统作为广泛临床领域的宝贵辅助手段具有相当大的前景。然而，它们的部署同时引起了重大的伦理问题，如果不加以解决，可能会破坏医患关系的信任、互惠和整体质量（表 1）。因此，必须在基于人工智能的技术的利用与维持医疗专业人员与其患者之间人性化和负责任的关系之间取得合理的平衡。为了实现这种平衡，必须始终坚持尊重患者及其权利的基本原则，同时也要履行尽我们所能履行职责的职业义务。

表 1. 人工智能在医疗保健中的主要应用及其对治疗关系的影响

人工智能在医疗保健中的应用	机会	主要伦理问题
辅助诊断（放射学、皮肤病学、数字病理学）	提高诊断准确性;减少错误;更快的诊断	临床权力从从业者转移到算法;与人工智能相关的错误;去技能化;对知情同意的质疑;决策过程的不透明性
预测医学和大数据分析	个性化预防;改善护理途径	过度的患者分析;污名化的风险;数据机密性问题
生成式人工智能（报告、临床文档、通信）	更快、更及时地起草报告和记录，从而提高效率	错误或误导性信息的风险;失去机密性
远程医疗和自动分诊	更广泛、更快速地获得护理;远程监控	减少人际互动;排除数字素养较低的患者;过度依赖自动化系统的风险
临床决策支持系统（CDSS）	更高的治疗精度;减少处方错误;促进量身定制的个性化治疗形式	临床权力从从业者转移到算法;与人工智能相关的错误;去技能化;对知情同意的质疑;决策过程的不透明性
手术和辅助机器人	更高的手术精度;减少侵入性;支持日常护理活动	与患者的身体和情感距离;高成本和不公平的获取;去技能化

本文旨在概述与人工智能技术相关的医患关系的潜在挑战。这样描述的情景有望成为定义未来研究重点和循证实践的宝贵框架，以保护和加强治疗联盟。任何不承认和尊重均衡标准和界限的行为都可能损害护理的质量和有效性，从而危及患者的健康和福祉。

2. 方法

本研究是对科学文献和最重要的国际政策文件的叙述性回顾。选择这种方法是为了突出生物伦理学、医学和监管的观点，通过系统评价的严格定量标准来全面概述这些观点可能具有挑战性。借用了 PubMed、Scopus、Web of Science 和 Google Scholar 数据库，以及世界卫生组织 (WHO)、联合国教育、科学及文化组织 (UNESCO) 和世界医学协会 (WMA) 发布的官方文件、建议和政策文件。这些机构的选择是凭借其权威，因此，它们有能力影响各个国家未来的监管选择。为了评估这些国际机构的建议是否已纳入欧盟、中国和美国，还考虑了这些国家（或欧盟等超国家机构）的主要监管框架。欧盟、中国和美国被选为世界三个主要地区的代表性司法管辖区。该分析包括 2015 年至 2025 年间发表的学术文章、指南和机构报告，重点关注人工智能的临床应用以及伦理和法律方面（知情同意、信任、自主、公平、尊严）。这种方法能够确定四个主要主题领域：为患者使用人工智能技术的理由；与患者沟通人工智能的使用情况；保密性；以及治疗联盟和医疗保健专业精神。作者排除了以英语或意大利语以外的语言发表的研究，以及那些以技术工程为主但缺乏伦理或法律分析的研究。

3. 结果

3.1. 患者使用人工智能技术的原因

2021 年 11 月 23 日，教科文组织通过了一项涉及不同应用领域的人工智能伦理问题的重要建议书，并将其作为其工作任务的一部分。该文件详细阐述了为正在研究的调查提供信息的许多要点，值得注意的是，它明确提到有必要投资于探索人工智能系统对人际关系影响的研究，以及确保对患者与医护人员关系的重要性给予应有的关注。

在涉及的伦理问题中，该文件指出，人工智能系统的使用必须得到以下论点的支持：（a）所选择的人工智能方法应适当、可取和相称，以实现合法目标；（b）所选择的人工智能方法不应侵犯或滥用人权，也不应侵犯基本价值观；（c）所选择的人工智能方法应适合上下文并基于严格的科学原理。

在这一点上，强烈建议在决定是否在特定环境中实施人工智能技术时采取谨慎的态度，同时考虑到对人类尊严的尊重，优先考虑患者的福祉和健康，而不仅仅是基于类似行业的成本效益分析。例如，值得考虑的算法系统旨在预测重病患者的预期寿命，识别那些不会从住院中受益并因此应该呆在家里的人。该公司为描述决策过程而采用的基本原理可能未知，但这样的系统可能会主要根据隐藏的经济利益来评估人类及其预期寿命，从而有可能侵犯患者的人格尊严和自主权。

世卫组织于 2021 年 6 月制定的关于“人工智能促进健康的伦理和治理”的指导文件中更广泛地讨论了这些方面和复杂性。本报告的前提是需要超越过度依赖人工智能技术作为灵丹妙药的方法，因为基于人工智能的解决方案也不能免受偏见和错误的影响。

具体来说，任何人工智能工具的可靠性都取决于为生成临床决策而使用的数据的质量和数量。如果数据不适合提供有关具体病例特征的可靠指征，则人工智能工具很可能在诊断和治疗中出现错误。如果有强有力的临床证据表明人工智能系统在特定任务上优于人类系统，那么使用该系统做出精确、明确定义的决策可能是完全合理的。如果将判断留给人类，而机器人可以更快、更准确、更具体地完成判断，一些患者可能会经历可预防的发病率和死亡。如果没有机会抵消收益，则尤其如此。然而，在缺乏此类证据的情况下，或者如果被认为没有定论，依赖人工智能可能会对患者护理的质量和安全性产生不利影响，从而侵蚀护理关系的基础和潜在信任。

从圣托马斯·阿奎那的断言开始，即有用的东西本身并不具有（即，只要它是有用的）善的理由，而是作为善的手段的理由（就像苦药在有利于恢复的情况下是有用的一样），因此不能说什么是有用的，就其本身而言，必然是好的。相反，善的东西也真正有用。这应该引导我们反思对人工智能技术采取适当评估方法的必要性，与其说是在特定情况下追求特定利益的有利策略，不如说是作为有效专业实践的一般条件的基本要素。

因此，创建持久且保护隐私的数据访问机制，以促进使用高质量数据更好地训练和验证人工智能模型，将使人工智能更加安全。

因此，文献中普遍认为，有必要防止人工智能工具的设计或编程方式可能通过优先考虑经济利益来侵犯患者的健康权、自决权和保密权。从这个角度来看，人工智能不应改变医患关系，这种关系传统上以保护医生责任下的患者为导向。

3.2. 使用人工智能与患者沟通

医院或医疗保健提供者是否会向患者披露人工智能在决策过程中被用于支持、验证甚至取代医生，这是值得怀疑的。在将技术用于治疗或诊断目的之前获得患者许可的先例是没有的。然而，知情同意的基础和更广泛的公众对医疗保健的信心可能会因人工智能在医学中的使用及其使用的遗漏而受到质疑。这个问题源于人工智能在决策和临床治疗中的应用是否会损害获得知情同意的任何理由的问题，包括保护、自主、避免滥用行为、信任、自我所有权、非支配性和个人诚信。医生应以公开、直接的方式及时告知患者人工智能的使用。

继续上述示例，如果人工智能系统被配置为根据住院费用数据建议护理途径，并且医生在没有进行批判性评估的情况下接受其关于生活机会分配的预测，则决策过程将受到不利影响，导致患者整体待遇不公平。虽然患者最终在决策过程中拥有权力，但可用选项的呈现方式可能会影响他们的决策，对他们在医疗保健选择中的自主权和自决权产生重大影响。

医生必须尽一切努力向患者阐明使用人工智能背后的基本原理、其机制及其可解释性，即人类能够掌握模型决策或预测背后原因的程度，这有助于使复杂人工智能系统的“黑匣子”更加透明和易于理解，并使用户能够理解给定模型为何得出特定结果。因此，医疗保健专业人员应彻底阐述收集和依赖的数据类型，以及如何利用这些数据并与外部各方共享，并详细说明为维护患者隐私而采取的安全措施。医生应该对使用人工智能技术可能引起的任何偏见、隐私问题或数据泄露开诚布公。从长远来看，人工智能在医疗保健和健康科学（包括临床试验和医院实践）中的使用只有通过确保透明度、可解释性和可理解性才能取得成功。所有这些组成部分都是有效的，因为它们代表了建立信任的基础，这对于推进人工智能在医学中的有效应用至关重要（第 26 页）。此外，正是人工智能决策的可解释性使医生能够评估其准确性并将其用作“第二意见”。因此，如果人工智能工具要充当决策支持系统而不是替代医生的决策角色，那么可解释性至关重要。在这种情况下，利害攸关的不仅是职业道德和道德坚持性，而且是支撑护理本质的更深层次的方面，这不能简化为单纯的技术-科学互动，不可避免地带有去人格化的色彩。

在决策受到人工智能算法影响或基于人工智能算法的情况下，必须向患者提供

全面的信息，特别是在此类决策可能影响其安全或人权的情况下。在这种情况下，个人还应有权要求负责监督此类技术及其实施的相关人工智能实体或公共部门机构进行澄清。

此外，至关重要的是，个人有能力理解支持每项可能影响其权利和自决权的决定的基本原理。此外，应允许他们向公共或私营部门组织的指定雇员提交意见，该雇员将负责审查和可能修改该决定。在商品或服务直接交付给客户或在人工智能系统的帮助下的情况下，人工智能行为者有责任确保及时、适当地通知用户。

因此，必须告知患者他们正在与人工智能系统互动，因为这可以促进与医疗保健专业人员和知情同意机构建立基于信任的关系，并防止医疗保健中新颖和更阴险的家长作风形式的出现，这可能植根于“计算机最了解”的未经证实的说法。

因此，人工智能在临床实践中的使用似乎使医患关系复杂化，因为它要求医生告知患者采用此类工具，并在可能的情况下解释其功能和可靠性程度。因此，原则上，用于治疗遭遇的人性维度的时间应该增加。

3.3. 保密

在医生和患者关系中维护隐私权的最重要价值观之一是保密。自希波克拉底时代以来，它在医学领域的重要性就得到了认可。

保护数据无异于维护信任。当患者怀疑他们的健康信息可能超出严格的护理目的时，他们可能会隐瞒相关细节。这种隐瞒不仅破坏了治疗联盟，还削弱了医生提供准确诊断和有效治疗的能力，从而危及整体护理质量。

关于人工智能系统在医疗保健领域的巨大进步，欧洲委员会生物医学与健康领域人权指导委员会于 2021 年 12 月委托编写的最新报告强调了不受限制的创新如何在这方面构成重大威胁（第 55 页）。

具体来说，当人工智能系统在医疗保健领域越来越频繁地被开发、使用和依赖时，可能会有更大的要求来生成或选择高质量的真实世界患者数据集来进行系统测试和培训。在这方面，正如几项研究也指出的那样，创新可以通过两种方式损害机密性和隐私。首先，允许第三方访问（去识别化的）患者数据和电子健康记录，以测试和构建人工智能系统。其次，人们担心医生可能倾向于要求额外的测试和分析，以训练或评估人工智能系统，而不是为了他们的临床利益。这个问题尤其重要，因

为它使患者面临与数据泄露或其他违反保密规定相关的不必要风险，以及不断增加的医疗费用。任何具有可疑治疗价值或明显仅由其对人工智能系统测试或开发的有用性驱动的数据生成都可能侵犯隐私权。

将患者健康记录用于测试和训练人工智能系统至少应遵守足够的去标识化和隐私增强协议。其中一种选择是差分隐私，它涉及添加噪声以防止识别数据集中的特定个体。

世卫组织的指导文件也表达了类似的担忧，该文件建议制定明确、更全面的数据管理计划，并强调仅靠保密义务可能不足以保护用于人工智能健康技术的数据。

鉴于医疗监测的潜在需求，联合国教育、科学及文化组织（UNESCO）敦促会员国在监管人工智能应用中的医疗保健预测、检测和治疗解决方案时要特别小心，此外还要确保遵守所有适用的国家和国际数据保护要求，维护隐私保护措施，并建立有效的机制来确保个人数据的个人数据被分析者被告知其使用并给予同意。此外，这一过程绝不应妨碍个人获得医疗保健。

保密不仅是一项法律义务，也是一项道德和哲学核心价值观，这是医疗保健本质的重要组成部分。因此，医学中的人工智能绝不能放弃对保密性的严格尊重，以防止患者失去对其信息的控制、对医疗保健专业人员的信任，并最终失去被视为人类而不是单纯统计数据的眼光。不遵守保密价值可能会给患者带来疏远感，他们可能会认为自己在一个不再承认他们的个性，而是将他们完全视为健康数据的存储库或来源的系统中的“资产”。

因此，即使在保密问题上，人工智能工具也可能带来重大风险。在这方面，无论是否存在具体的国家法规，医疗保健专业人员都必须恪守保密义务，坚持患者福利的首要地位，确保敏感健康信息的保护优先于人工智能开发者和生产者的商业利益。

3.4. 治疗联盟和医疗保健专业精神

许多当代研究证实了欧洲委员会报告的调查结果，表达了对医疗专业人员过度依赖人工智能（他们将患者护理的完全控制权委托给该技术）可能导致对此类技术的不适当依赖。这种过度依赖可能导致医生技能下降（技能下降），最终降低其诊断敏感性。

当医生和患者都能够为最终选择做出贡献（医生通过提供医学专业知识和同理心，患者通过提供信息、价值观和个人偏好）时，即使人工智能参与了这一过程，共享决策（SDM）也是真正实现的。将医生和患者排除在决策过程之外可能会损害医生的职业自主权，因此，技术解决方案的广泛实施可能导致失去控制和被排除在对话之外的风险，而对话可以说是医疗保健专业人员 and 患者之间的临床护理和共同决策至关重要的。此外，如果人工智能技术最终缩小了医生与患者之间的沟通和互动范围，它可能会限制前者为患者提供最有效治疗干预的能力，从而损害一般支持性护理的原则，其中包括在患者通常最脆弱时人际互动的好处。因此，人工智能有可能进一步使医疗实践非人化并导致家长式回归，但这一次是人工智能而不是医生应用规则。例如，自动分诊系统扩大了获得医疗保健的机会，特别是对于服务不足的人群和资源有限的社区。然而，技术中介存在降低人际接触的即时性的风险，从而限制了感知非语言线索、情感和隐性需求的能力，并可能削弱治疗联盟，并侵蚀支撑医患关系的关系信任。生物伦理学学术强调，人的尊严不能被简化为单纯的技术参数。世界医学会同样强调医患关系在医疗保健人工智能系统的开发和实施中的重要性。

此外，随着人工智能生成的医疗意见变得越来越广泛且值得信赖，患者可能会转向人工智能软件来获得医疗保健指导。因此，人工智能生成的医学意见的出现可能会导致一直拥有所有知识和经验的医生与患者或家属之间的关系进一步发生变化，患者或家属现在可以通过谷歌的 Bard 或 OpenAI 的 GPT-4 等大型语言模型（LLM）聊天机器人访问越来越复杂和准确的人工智能生成的意见。GlassAI 是专为医疗领域设计的消费技术。它基于 GPT-4 算法，允许用户进入临床场景，生成鉴别诊断或临床管理方案。随着此类技术的普及和普及，患者可能会对临床检查产生一定的期望。

另一方面，生成式人工智能工具越来越多地用于起草医疗报告、综合临床文件以及使患者更容易理解医学语言。通过这样做，它们使医生能够更有效地执行任务，从而腾出更多时间来倾听患者的意见，全面评估他们的需求，并提出治疗方案和预测结果，因为人工智能有助于数据检索和解释。也就是说，医生可以花更多的时间与患者接触，而不是关心与电子健康记录相关的耗时的官僚职责和“文书工作”，众所周知，电子健康记录会造成不满和疲惫。

除了通过转录和记录过程节省时间的潜力外，技术还可能有助于减少医生的倦

息并与医学的人文方面建立更深入的联系。通过能够以易于理解的格式及时访问患者信息，这种方法还有助于患者护理的民主化。

人工智能技术可以促进接触和沟通，从而加强医患之间的联系，而不是破坏它。例如，他们可以分析各种治疗方案，以便医生可以通过披露益处和风险来与患者讨论它们。它们还可以个性化信息，使患者更容易获得信息，在某些情况下比常用的临床指南更有效。

科学家们正在进一步突破界限。未来，已经在商业领域成功测试的工具——能够通过面部表情以及言语和写作模式来识别客户的情绪——可能会在医疗保健领域得到应用，以提供越来越个性化的服务。帮助了解患者的情绪状态对于及时解决不满情绪非常有价值，如果不加以控制，可能会破坏信任关系。然而，将人性层面的护理委托给机器，即使是最先进的机器，也是一个严重的错误。事实上，人工智能无法提供人际接触，而人与人之间的接触通常被视为表现出兴趣和同理心的一种方式，甚至提高医生的诊断技能。

在医学技术维度方面，手段与目的之比的潜在倒置可能会导致医患关系发生负向转变。如果医生被视为仅仅是数据的处理者，而不是护理提供者，并且患者被简化为一个统计数据，而不是一个人，那么医疗决策最终可能会失去其对话价值，因为人工智能提供了基于大数据的最佳解决方案，从而减少了个性化的余地。必须承认医生和技术人员之间的区别，并强调这些数字永远不能互换。医生不应受制于机器人或人工智能技术的指令，进一步加剧了这种区别。医患关系和联盟应该加强，而不是削弱；然而，正如文献所表明的那样，人工智能不仅会带来好处，还会在医患关系的各个方面带来风险（表 2）。

表 2. 人工智能集成如何影响医患关系的不同层面

维度	潜在好处	道德风险和关键问题
护理质量	更高的诊断和治疗准确性;个性化治疗;预测医学	由于数据偏差而可能出现的错误;人工智能系统的不透明度;与人工智能相关的错误;专业去技能化;优先考虑患者利益以外的计划选择
可及性	缩短等待时间;提高护理的连续性;远程跟进的可能性	排除数字素养较低的患者;由于经济或技术障碍导致的获取不平等
医生自主权	决策支持;减少官僚工作量;更多时间与患者互动	专业去技能化;决策权从临床医生向人工智能系统转移

维度	潜在好处	道德风险和关键问题
知情同意和信任	有更多时间通知患者	信息处理的复杂性增加;人工智能系统的不透明度;对医生失去信任
公平	通过标准化减少差异的潜力	人工智能对少数群体的用处有限,因为算法是在不能充分代表他们的数据集上训练的
数据保密	自动匿名化	未经同意使用数据;违反匿名性

4. 讨论

通过实施人工智能来改善医疗保健领域的潜力是无庸置疑的。例如,基于远程医疗的监测,对患有不同性质的严重疾病的患者,显示出高度的可靠性和显著的益处。此外,在医生与患者互动时自动填写电子病历的工具可以提高医疗咨询的效率和效果。人们对能够诊断和开具治疗处方的人工智能工具的使用表示了担忧,因为此类技术有可能取代医生。此外,由于人工智能固有的易出错性,联合国教科文组织、世界卫生组织和 CDBIO 等著名组织以及循证研究已经达成共识,即医生应始终保留监督人工智能正常运行和控制决策过程的义务。因此,简单地假设人工智能可以节省医生的时间是错误的,而可以合理地得出结论,在医疗保健领域实施这些技术需要一种更关键的方法,对批判性思维和推理进行更有针对性的投资。此外,鉴于人工智能将对每个应用其医疗领域带来独特的复杂性,还强调了针对特定医疗领域进一步定制人工智能风险评估的必要性。风险评估应进行定制和定制,以匹配特定领域的显著特征,因为临床、社会和伦理风险和限制因放射学、外科、基因组学、心理健康、儿童健康和家庭护理等领域而异。例如,在放射学领域,来自欧洲和北美的几个主要放射学协会(包括美国放射学会、欧洲放射学会、北美放射学会、医学成像信息学会、欧洲医学成像信息学学会、加拿大放射科医生协会和美国医学物理学家协会)联合发表声明解决在放射学中使用人工智能相关的伦理困境。所有这些协会都指出,放射科学界应立即开始制定人工智能道德和实践准则,鼓励任何有利于患者和更大利益的应用。随着其他科学协会技术的进一步发展,也应该开展类似的努力并不断更新,以便优先考虑合乎道德和公平的实施和应用。

为了验证任何基于人工智能的决策的有效性,临床医生必须对之前由人工智能分析过的相同仪器测试和临床数据进行评估。也就是说,为了检查诊断和治疗的正

确性，或人工智能提供的任何其他贡献，临床医生还必须做出诊断并决定治疗。

生成式人工智能在诊断和治疗决策中的使用似乎使医疗实践和学生培训更加复杂，因为医学课程应该更加重视人工智能素养，以增强患者对医生使用的工具的理解，从而培养基于信任的关系。除了与患者关系中的常用信息外，医生还必须解释所选人工智能工具的工作原理、可靠性以及其使用背后的逻辑和动机，以及与保护所涉及的健康数据有关。这项额外的信息工作是一项职责，也是一项非常复杂的任务，这反过来又需要与其他执行官僚主义和重复性活动的人工智能工具来平衡。

医患沟通的过程不仅仅是信息的传达，还是一种建立在信任、同理心和相互理解基础上的关系和解释。人工智能的整合有可能以多种方式影响这种微妙的平衡，导致沟通的不透明度和不对称性增加，特别是如果不是所有参与者都能访问相同的信息并具有相同的理解水平。由于缺乏或缺乏可理解的解释，这可能会导致患者不信任，从而可能破坏自主感，甚至降低他们对治疗周期的依从性。这种担忧表明了德国哲学家汉斯·约纳斯（Hans Jonas）提出的观点，他警告说，先进技术可能会产生潜在的意外后果，并认为此类创新可能会自相矛盾地导致更高的不确定性。因此，乔纳斯指出需要扩大责任和意识。

如表 3 所示，世卫组织、教科文组织和世界监测协会提出的上述建议尽管具有权威，但尚未在全球层面得到统一实施。

在欧洲，《通用数据保护条例》（GDPR）要求数据收集仅限于绝对必要的范围（数据最小化），仅按规定目的按比例使用（相称性），并在责任和可追溯性（问责制）的框架内进行处理，这需要记录此类数据的使用方式和原因。具体到人工智能，GDPR 第 22 条规定了不受仅基于自动化处理的决策的权利。

此外，欧洲法规 2024/1689（人工智能法案）第 14-15 条将健康相关系统归类为“高风险”，从而对技术稳健性、透明度、可解释性（为算法决策提供可理解原因的能力）和人工监督提出了要求，正是为了确保患者和医生都保留对算法驱动决策的最终控制权。

同样重要的是法规 2017/745(附件 I, 第 23 节)，该法规要求医疗软件制造商向医疗保健专业人员提供明确的说明，从而使他们能够为患者提供有关所用人工智能工具的充分信息。

中国采取了完全不同的方法。《个人信息保护法》第 13 条规定了包括健康数据

在内的敏感个人数据的使用，并将同意确立为法律依据。然而，法律并没有确切规定必须向患者提供什么样的信息：不强制详细解释算法是如何运作的；关于在处理个人数据时使用技术工具的一般性通知被认为是足够的。此外，第十三条规定，在履行国家义务或义务、应对突发公共卫生事件或保障自然人的生命、健康或财产安全所必需的情况下，无需征得同意。

国家药品监督管理局（NMPA）2022 年基于人工智能的医疗器械指南涉及技术要求、临床验证和数据质量。该文件主要关注制造商与公共当局之间的关系，而不是医患沟通。事实上，NMPA 法规要求医生了解该算法的局限性和性能，但医生没有义务以清晰易懂的方式将这些信息传达给患者。此外，目前的监管框架没有编纂任何条款，支持患者对算法决定提出质疑、拒绝在治疗环境中使用或要求进一步澄清的权利。因此，上述指南并未为患者建立与《欧洲人工智能法案》中规定的具有约束力的透明度义务相媲美。

中间模式以美国为代表，与欧洲和中国不同，美国不存在联邦监管框架。在美国，适用的法规也规定，人工智能可以支持（但永远不会取代）医生的决策角色。

然而，美国的患者无权要求对算法决策进行人工审查，正如《欧洲人工智能法案》所规定的那样。这并不排除患者一方面拒绝基于算法提出的治疗的能力，另一方面，在通过人工智能确定诊断或治疗的情况下，向患者隐瞒此类信息可能构成违反医生披露义务的可能性，从而产生损害赔偿责任。

根据美国医学会（临床实践中最著名的旗手之一）的说法，告知患者人工智能使用情况的义务与此类技术可能对患者安全构成的风险水平成正比。一些学者指出，除了伤害风险外，还应进一步指导在仅告知患者人工智能工具的使用、征求他们的明确同意或省略信息和同意之间的选择：即患者能够行使决策权的程度。彻底知情同意的核心是咨询，咨询受到此类创新实践所带来的潜在和新颖的独特特征的影响。因此，这种额外的复杂性对于医疗保健领域的其他革命性技术和方法来说并不奇怪，并且与整个表观遗传学、基因编辑和个性化/精准医疗，所有这些领域也必然会受到人工智能、机器学习和大数据处理的影响。因此，咨询旨在成为一个有机和多层次的过程，而不仅仅是一组咨询，以使患者充分意识到并能够理解和评估风险、益处以及他们的选择对他们生活的独特影响。当伤害风险和患者维护其意愿的能力都很高时，强烈建议征得同意。相反，当伤害风险较低且没有现实的前景表明披露会影

响患者的选择时，医生无需提供此类信息。

表 3. 欧盟、中国和美国的监管比较

维度	欧盟	中国	美国
主要监管来源	-GDPR2016/679 -《2024/1689 年人工智能法案》 -第 2017/745 号条例	-个人信息保护法（个人信息保护法，2021） -NMPA 指南（2022）	-FDA，临床决策支持指南（2022） -FDA、AI/MLSaMD 行动计划（2021 年） -缺乏统一的联邦框架
人工智能与医生的关系	人工智能不能替代人类医疗保健专业人员，但它应被视为一种辅助工具	人工智能不能替代人类医疗保健专业人员，而是一种辅助工具	人工智能不能取代人类医疗保健专业人员，但只能作为辅助工具
对患者的透明度	-当人工智能影响诊断或治疗时通知患者的义务 -可解释性 -被告知局限性和误差范围的权利	-没有义务披露人工智能的使用或解释算法的功能	-AMA（2024）：与危害风险水平成正比的披露
对算法决策提出异议的权利	-人为干预和拒绝自动决策的权利（GDPR 第 22 条）	-不提供	-不承认医生的权利，尽管后者可能因未能披露人工智能的使用而承担责任
医患关系的概念	-保障导向模式：患者自主权至上	-集体主义模式：信任医生/机构；有限知情同意 -国家和技术控制的中心性	-中间模型：通过知情同意和医生的法律责任来保障自主权

5. 结论

医学行业和医患关系的未来变化将取决于哪些类型的人工智能进入临床实践。如前所述，如果人工智能取代医生从事行政性和重复性任务，他们将有更多时间陪伴患者和正确管理人工智能系统。相反，如果除了行政职责之外，医生还需要了解人工智能如何生成诊断或治疗以验证其合理性，那么医患关系可能会受到负面影响。

或者，他们可能会完全依赖人工智能。然而，在这种情况下，医生将失去专业技能和自主权。世卫组织明确表达了这种担忧，他们指出“如果人工智能的引入没有得到有效管理，医生可能会不满意，甚至失去医疗实践”（第 61 页）。

当然，随着技术的发展，可以肯定地说，人工智能和机器学习距离彻底发挥其潜力还很遥远，它们的边界现在甚至无法定义。事实上，随着量子计算研究的推进，它导致了量子机器学习的出现，它代表了量子计算和人工智能的交叉点。这种融合

有望解决目前即使是当今最强大的“传统”系统也无法解决的计算问题。这一发展的潜力是巨大的，可能为日益强大和影响深远的应用铺平道路。因此，目前不可能进行遥远的结论性风险收益分析，也不可能对即将出现的问题和困境进行明确的评估。

然而，根据本文概述的分析，只要医学界应保持目前将人工智能视为日常临床实践中宝贵资产的概念，并由合格且经过充分培训的医疗保健专业人员监督，这些变化可以说会产生积极影响。这种方法应被视为确保维护医患关系至关重要。人工智能在医疗保健中的作用绝不应被视为医疗专业人员的替代品或替代品。相反，它应该被视为一种强大的手段，可以最终达到提高诊断准确性、优化临床流程以及腾出时间来处理确实不可替代和不可协商的事情：与患者的人际关系。正如汉斯·乔纳斯提醒我们的那样，技术必须负责任，其在医学中的应用必须加强而不是破坏护理伦理。

译文二：

医学伦理中的人机协议：基于患者自主权的大型语言模型案例评估

Vamshi Mugu, Brendan Carr, Ashish Khandelwall, Mike Olson¹, John Schupbach, John Zietlow, T N Diem Vu, Alex Chan¹, Christopher Collura, John Schmitz¹, 徐健（译）

1. 介绍

美国医学学会将医学伦理学定义为临床医学实践的道德框架。由 Beauchamp 和 Childress 提出，这四项原则，即仁慈、非恶意、患者自主和正义，通常被视为当今该领域的教学和实践医学伦理学的基石。在这四项原则中，患者自主权占有特别重要的地位，因为它在由于四项原则之间的冲突而出现伦理困境时经常优先于其他三项原则。

大型语言模型（LLM）代表了人工智能的新范式，具有处理和生成自然语言文本的卓越能力。基础法学硕士可以执行许多不同的任务，但有时缺乏特定领域的功能。尽管基础法学硕士的能力已转化为医学教育、研究和实践中的众多应用，人们仍然担心其准确性、偏见和潜在的滥用。解决这些问题的拟议方法中反复出现的主题是人类专家参与法学硕士评估和治理的重要性。本研究也强调了人为因素的重要性，促使各大洲各管理机构制定了指导方针。

关于法学硕士在医学伦理学中的作用的文献明显稀缺，特别是关于如何提高基础法学硕士在医学伦理学特定任务上的可靠性的文献。医学伦理学中法学硕士的比较评估几乎不存在，只有 ChatGPT 得到了相当程度的评估。尽管知识体系，特别是提示工程，为了从法学硕士那里获得更好的反应，总体上正在扩大，缺乏在医学伦理学背景下对这些技术的讨论。我们冒险进行了一项大胆的事业，不仅评估基础法学硕士在患者自主权中分析病例的能力，而且还探索提高法学硕士响应可靠性的技术，从而试图增加对该技术的信任，同时强调人为因素的不可或缺的作用。

2. 方法

2.1 伦理考虑

在获得妙佑医疗国际机构审查委员会（根据批准 25-001042 管理研究伦理）的批准后，撰写了 44 个需要回答是或否的患者自主权的假设病例。本研究没有使用真实的患者信息。第一作者从文献以及来自个人和集体临床经验，重点关注同意能力、职业暴露、保密性、未成年患者的知情同意、患者偏好、治疗拒绝和培训需求。

2.2 假设案例

一个假设的案例在文本框 1。法学硕士和小组中的医生（后文描述）对病例的来源和其他人的反应视而不见。

文本框 1 假设案例示例。请注意，假设的案例涉及同意和拒绝治疗的能力。

假设案例示例

大卫是一名 79 岁的男性，有心肌梗塞病史，他的妻子玛丽因大卫的胸痛拨打了 911。疼痛始于 30 分钟前，与他之前需要住院的一些发作相似。紧急医疗服务（EMS）团队到达后，大卫拒绝任何干预。大卫似乎因疼痛而苦恼，但却有想法。EMS 团队应该带 David 去最近的医院吗？

本研究选择了三个基础法学硕士：ChatGPT 版本 4o（ChatGPT）、LLaMA3.170bBF16（LLaMA）和 Gemini1.5（Gemini）。尽管开源 LLaMA 在本地与 Ollama 版本 0.1.33（Ollama, Inc）一起使用，但 ChatGPT 和 Gemini 这两个闭源模型是使用各自的 Web 界面访问的。在适用的情况下，使用默认参数（例如，温度为 0.8，top_pOllama 为 0.95）。由于案件的假设性质，没有担心不遵守《健康保险流通与责任法案》（HIPAA）。五名医生（P1-P5）分别获得了急诊医学、外科和放射学的委员会认证，组成了医生小组。

2.3 学习规划

我们的研究分为两个阶段（图 1）：评估阶段和改进阶段。评估阶段旨在比较基

基础法学硕士的反应与医生小组的反应，而改进阶段旨在优化法学硕士的反应以提高与医生小组的一致性。法学硕士和小组中的医生对彼此的回答不知情。使用 Fleiss κ 计算小组医生和基础法学硕士之间的观察者间一致性。Cohen κ 用于将每个 LLM 的回答与医生共识进行比较，医生共识定义为小组中医生的大多数反应。出于统计分析目的，LLM 中的缺失值被视为“否”。为了说明这一点，成对协议被描述为热图，并以表格格式列为比例。Python 统计模型库（版本 0.15.0）用于统计测试。与 $\kappa < 0$ 的一致性被认为是较差的，0 到 0.2 被认为是轻微的，0.21 到 0.4 被认为是一般的，0.41 到 0.6 的一致性是中等的，0.61 到 0.8 是实质性的，0.81 比 1 是几乎完美的。尽管我们保留了几种在改进阶段采用的提示工程技术，但我们仍然在评估阶段使用了一些技术，例如角色扮演、强力和思维链。

在改进阶段，使用思维链、N-shot 提示、定向刺激、版本控制、改写和响应以及长上下文提示等提示工程技术来改进 LLM 响应。这个迭代过程的目标是改进 LLM，以产生更符合医生共识的响应，从而减少任何统计学上的显著差异。迭代过程的停止点被确定为无法再减少不同响应数的时候。使用 McNemar 检验评估改进程度。小于 0.05 的 P 值被认为是显著的。

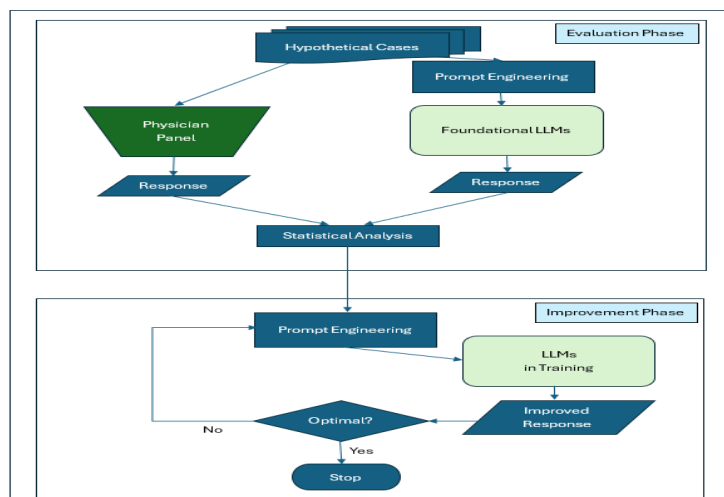


图 1. 两阶段设计，包括评估阶段和改进阶段。评估阶段将基础法学硕士的反应与医生小组共识进行比较。改进阶段旨在提高法学硕士与医师小组共识之间的一致性。LLM：大型语言模型。

示例人模交互

在评估阶段和改进阶段提供了示例人模交互文本框 2。这个假设的案例涉及未成年患者的同意，并展示了评估和改进阶段的人模交互。本示例说明了两种提示工程

技术的使用，即问题细化和思维链。

文本框 2. 在评估和改进阶段进行人模交互的假设案例示例。这里演示了两种提示工程技术，即问题细化和思维链。

示例与演示评估和改进阶段的模型的交互

假设案例：雷切尔是一名来自弗吉尼亚州的 9 岁女孩。她在学校篮球场摔倒，被老师送往急诊室。她被发现肘部骨折，需要非紧急手术固定。外科医生可以进行手术吗？

评估阶段的交互片段

双子座的回应

答：是的

思考过程：

1. 外科医生有责任以患者的最大利益行事。在这种情况下，手术是必要的，以治疗肘部骨折和预防并发症。
2. 延迟手术可能会导致病情恶化、疼痛和潜在的长期残疾。
3. 因此，进行手术符合患者的最佳利益。

改进阶段的交互片段

精炼问题(由 ChatGPT 应人类要求提供)：鉴于来自弗吉尼亚州的 9 岁女孩 Rachel 在学校篮球场摔倒，被老师带到急诊室，在那里她被诊断出患有肘部骨折，需要非紧急手术固定，外科医生是否拥有进行手术的必要同意和权限？

双子座的回应

答：没有

思考过程：

1. 雷切尔是未成年人，因此缺乏为手术提供知情同意的能力。
2. 她的法定监护人，可能是她的父母，必须提供同意。

3. 结果

在 44 例病例中，6 例侧重于知情同意，6 例侧重于同意能力，6 例侧重于职业暴露，6 例侧重于培训需求，7 例侧重于拒绝治疗，6 例侧重于患者偏好，7 例侧重于保密性。一个假设案例示例在文本框 1. 小组中的医生之间基本达成一致($\kappa=0.77$)。基础法学硕士之间也存在基本一致性($\kappa=0.73$)。基础 Gemini 与医生共识略有一致($\kappa=0$)，基础 LLaMA 与医生共识略有一致($\kappa=0.14$)，基础 ChatGPT 与医生共识略有一致($\kappa=0.32$)。成对协议描述在图 2 作为热图以供说明。但是，请注意，改进阶段旨在提高法学硕士与医生小组共识之间的一致性，而不是个别医生的反应。

在改进阶段结束时，所有改进的法学硕士与医生共识之间的一致性相当大或更高(ChatGPT 的 $\kappa=0.77$, 95%CI 0.59-0.96; 双子座的 $\kappa=0.73$, 95%CI 0.52-0.93; LLaMA 的 $\kappa=0.82$, 95%CI 0.65-0.99)。所有 LLM 的改善程度均具有统计学意义(ChatGPT 为

$P=.006$, Gemini 为 $P<.001$, LLaMA 为 $P<.001$)。改善程度列在表 1。

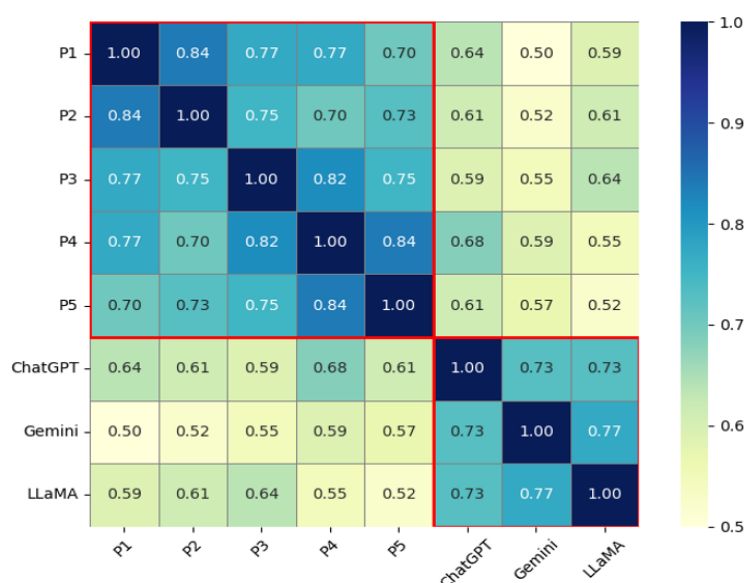


图 2. 成对一致性热图仅供说明之用。请注意，在评估阶段（用红框括起来）期间，大型语言模型和小组（P1-P5）上的医生之间存在基本一致性。然而，改进阶段（稍后讨论）旨在提高大型语言模型与医生共识的一致性，而不是与个别医生反应的一致性。

表 1. 改进前后每个大型语言模型与医生共识的一致性比例

类型	改善前 n	改进后 n	改善程度 (N=44) n (%)	P 值
ChatGPT	27	39	+12 (27)	.006
双子座	22	38	+16 (36)	<.001
骆马	26	40	+14 (32)	<.001

注：所有大型语言模型都经历了具有统计学意义的改进。

4. 讨论

4.1 主要发现

医学伦理学为临床医学实践提供了一个道德框架，特别是当它涉及医患互动背景下的医生责任和义务时。博尚和柴尔德里斯通常被认为为该框架奠定了基础，其中包含 4 项原则：仁慈、非恶意、患者自主和正义。尽管这些原则受到了广泛的批评、争论并辩护，尽管如此，它们仍然是现代医学伦理学教学和实践的中心。患者

自主原则在这四项原则中占有特别重要的地位，当由于 3 项原则之间的冲突而出现伦理困境时，患者自主原则通常优先于其他 4 项原则。它不断提醒医生，“患者必须始终处于他或她个人战斗的核心”和“患者的需求是第一位的”。

尽管法学硕士还处于起步阶段，但它们已经彻底改变了医疗保健的各个方面，从医学教育到实践和研究。法学硕士在曾经被认为与训练有素的人类密切相关的任务上表现出了非凡的准确性，例如在神经外科书面类似板状问题上优于医学生和神经外科住院医师。然而，法学硕士很容易受到限制，其中一些在医学上尤其令人担忧，例如捏造不基于证据的反应——这种行为被称为幻觉。。此外，对法学硕士等技术的依赖提出了医学伦理的基本问题，其中类人推理、情商以及对当地背景和价值观的理解至关重要。解决这些问题的拟议方法中反复出现的一个主题是人类专家参与法学硕士评估和治理的重要性。我们的研究也强调了人为因素的重要性，促使各大洲的各个管理机构制定了指导方针，并为有效交互铺平了道路，例如通过提示工程。提示工程领域正在迅速发展，已经采用了几种成功的方法，包括思维链、定向刺激、力量、自我纠正和一致性、角色扮演、反思、导轨引导输出和长上下文提示。

尽管关于人工智能在医学伦理学领域应用的文献很少，但有一些工作值得一提。在法学硕士的构思之前，Shalowitz 等人提出了一种基于人群的治疗指标，并声称其在为无行为能力患者做出的治疗决策方面优于传统的替代决策。基于人群的治疗指标因依赖统计证据而不是个体患者偏好而不尊重患者自主权而受到批评。利用法学硕士的自然语言处理能力，Earp 等人提出了 P4，一种个性化的患者偏好预测器，以解决这一特定问题。P4 的优越性主张是基于它纳入了先前患者治疗决策中的材料，从而创建一个“数字孪生”，以便在出现此类需求时代表无行为能力患者采取行动。

我们的研究可能与 Balas 等人对 GPT-4 的专家小组评估最相似。他创建了一组 8 个道德案例小插曲，并通过先验提示模板呈现给 LLM。然后评估法学硕士的回答，以评估推理的深度、考虑不同观点的能力以及对道德困境细微差别的敏感性。尽管我们的案例库要大得多，但也可以说更简单，因为我们不探索推理的深度。然而，与他们的研究最大的区别是在改进阶段。从最严格的研究设计意义上讲，他们的研究可以被认为是缺乏这个阶段的。通过不仅评估基础法学硕士，还探索提高其与医学伦理子集相关的表现的方法，我们冒险对这项有前途的技术建立一定的信心，同时提醒采用者人为因素的重要性。

4.2 局限性

这项研究受到一些限制。尽管比其他类似研究规模更大，但它仍然涉及一组有限的问题，涵盖相对较小的医学伦理部分。法学硕士在不同的临床场景或伦理主题下可能会有不同的表现。尽管我们采用了几种类型的提示工程，但没有使用其他改进技术，例如 RAG 和微调。随着时间的推移，工作无疑将继续提高基础法学硕士在医学伦理学方面的表现；多快以及是否会完全消除对人工监督的需求还有待观察。

5. 结论

在医学伦理学等领域使用基础法学硕士需要非常谨慎和人类专家的复杂参与。通过充分的测试并利用提示工程等成熟技术，法学硕士的表现也可以提高，即使在医学伦理学领域，类人推理、情商和情境感知至关重要。

***注：原文和译文版权分属作者和译者所有，若转载、引用或发表，请标明出处。**