多模态大模型在医疗领域的应用与治理研究

**摘要​**：  
多模态大模型（Multimodal Large Models, MLLMs）作为人工智能领域的前沿技术，通过整合文本、影像、信号等多源异构医疗数据，正在深刻改变医疗创新与临床实践的范式。本文系统综述了MLLMs在医疗领域的技术基础、应用进展与未来挑战，旨在为相关研究提供全面参考。研究表明，基于通用大模型（如GPT-4、LLaMA）的医疗领域微调（如LLaVA-Med、Med-PaLM M）已成为主流技术路径，通过跨模态对齐机制实现了影像-文本-基因数据的协同分析。在医疗创新方面，MLLMs已展现出显著价值：在智能诊断领域，多模态融合显著提升了肿瘤和罕见病的诊断精度；在药物研发中，跨组学数据分析加速了靶点发现与个性化治疗方案生成；在临床实践中，增强型临床决策支持系统（CDSS）有效优化了诊疗效率和标准化水平。然而，MLLMs的临床应用仍面临数据隐私与安全、模型可解释性缺失、算法偏见与泛化性不足、责任界定模糊等核心挑战。未来需要协同推进技术创新（如轻量化架构、联邦学习）、伦理治理（如偏见监管、透明性标准）与临床验证（如随机对照试验），才能实现MLLMs在医疗领域安全、公平、高效的应用。本综述为理解MLLMs在医疗领域的现状与未来发展方向提供了系统性视角。

****关键词****：多模态大模型；医疗人工智能；智能诊断；临床决策支持；医学影像分析；医疗数据融合；医疗伦理

1. 研究背景

当代医疗领域正面临信息爆炸、精准医疗需求骤增、效率瓶颈以及资源分布不均等多重挑战。人工智能（AI）技术在医疗领域的兴起与发展，为解决上述问题提供了全新路径。从早期的规则驱动系统到如今的数据驱动方法，AI逐步渗透至医疗的各个环节[[[1]](#endnote-0)]。特别是大模型（LLMs）的突破性进展，以其卓越的语言理解、生成与推理能力，开启了医疗创新的新篇章[[[2]](#endnote-1)]。与此同时，多模态学习的重要性日益凸显。医疗数据天然多源异构，包括文本、影像、信号、组学及临床记录等，其复杂性要求更高效的跨模态分析工具[[[3]](#endnote-2)]。在此背景下，多模态大模型（MLLMs）凭借整合跨模态信息的能力，展现出更全面理解复杂医疗场景的独特优势。本综述旨在系统梳理MLLMs在医疗创新与临床实践中的最新进展、关键应用、优势、挑战及未来方向。

1.1 当代医疗面临的挑战

随着全球医疗需求的增长，当代医疗体系面临诸多严峻挑战。一方面，医疗数据呈现出指数级增长，如何高效处理和利用这些数据成为亟待解决的问题[[2](" \l "anchor_2)]。另一方面，精准医疗对数据分析的深度与广度提出了更高要求，而传统医疗模式在效率和资源分配上的瓶颈愈发明显。此外，医疗服务资源分布不均的现状进一步加剧了全球医疗负担[[3](" \l "anchor_3)]。这些问题共同推动了人工智能技术在医疗领域的快速发展，并促使研究者探索更具潜力的解决方案。

1.2 人工智能与多模态大模型的潜力

人工智能技术在医疗领域的广泛应用，正在改变传统的医疗服务模式。尤其是大模型的突破性进展，使其在语言理解、生成与推理方面表现出强大的能力[1]。然而，单一模态的 AI 系统往往难以应对医疗数据的多样性。多模态学习作为一种能够融合多种模态数据的技术，其重要性逐渐显现。通过整合文本、图像、语音等多种类型的信息，多模态大模型（MLLMs）不仅提升了医疗数据的利用效率，还为医疗决策提供了更加全面的支持[3]。例如，在医学影像解读中，多模态大模型可以结合影像特征和患者的病史文本，更准确地判断疾病类型和严重程度[10]。这种融合能力在疾病诊断、个性化治疗等领域展示出巨大潜力[2]。

1.3 综述范围与结构

本综述聚焦于通用或医疗专用的大型预训练模型，特别是那些在医疗健康领域具有显著应用价值的多模态大模型。涵盖的应用领域包括但不限于诊断、治疗、药物研发、医疗管理等[[1](" \l "anchor_1)]。后续章节将围绕多模态大模型在医疗领域的技术突破、知识组织、模型构建与应用、对医疗决策的影响等方面展开详细探讨，以揭示其当前进展与未来发展方向[[2](" \l "anchor_2)]。

2. 多模态大模型在医疗领域的应用机遇与挑战

多模态大模型近年来在医疗领域的探索展现出前所未有的潜力，但其实际落地仍需克服一系列技术和伦理障碍。本章将从应用机遇和挑战两个维度展开分析，揭示这一技术如何在创新医疗服务模式、优化诊疗流程的背景下重新定义医疗健康生态系统。

2.1 应用机遇

多模态大模型通过整合不同类型的医疗数据模态（如文本、图像、视频等），为医疗领域带来了显著的应用机遇。韩普等（2023）[11] 提出了一种针对多模态医疗健康数据的知识组织模式，通过实体对齐和关系抽取构建知识单元，并在智能问答系统的实践中验证了该模式的有效性。这种方法不仅提升了多模态数据的处理效率，还为精准医疗提供了新的技术支持。例如，在智能问答系统中，该模式能够快速准确地回答患者关于疾病诊断和治疗方案的问题。​

此外，Clément Christophe 等人（2024）[12] 的研究进一步表明，大规模语言模型经过合理微调后能够适应复杂的医疗场景，例如患者病历分析和诊断辅助决策。这种基于参数高效微调的策略降低了资源消耗，同时保持了模型性能，为医疗 AI 的技术普惠化奠定了基础。例如，在分析患者病历文本时，微调后的模型能够快速提取关键信息，辅助医生进行诊断。​

另一方面，Yanis Labrak 等人（2024）[14] 开发的 BioMistral 模型展示了多模态大模型在跨语言医疗问答任务中的巨大潜力。通过自动翻译和多语言评估，该模型突破了语言障碍，为全球范围内的医疗知识共享和协作开辟了新路径。这不仅促进了医疗资源在全球范围内的公平分配，也为发展中国家提供了获取先进医学知识的机会。在此基础上，黄桂新等（2023）提出的基于多准则的医疗大数据智能化决策平台，结合 Hadoop 架构实现了多目标量化的医疗决策支持系统，进一步推动了多模态技术在管理复杂医疗决策中的应用。

2.2 应用挑战

尽管多模态大模型为医疗领域带来了诸多机遇，但其实际应用过程中仍面临诸多挑战。首先，在技术层面，如何高效整合异构数据并保证模型输出的可靠性是一个关键问题。韩普等（2023）[11] 指出，多模态数据的知识单元抽取过程需要依赖高质量的预处理和实体对齐，而这对于专业性强且格式复杂的医疗数据而言，存在较高的技术门槛。例如，不同医院的病历格式和影像数据标准各不相同，整合难度较大。此外，Clément Christophe 等人（2024）[12] 的研究也提到，大型语言模型的全参数微调在医疗领域中可能会导致过拟合或泛化能力不足的问题，而参数高效微调策略虽然能够缓解这些问题，但在极端场景下的表现仍有待验证。​

其次，伦理与隐私问题是多模态大模型在医疗应用中不可忽视的挑战之一。Yanis Labrak 等人（2024）[14] 的跨语言实验表明，当医疗问答任务涉及到敏感的患者信息时，模型的输出可能引发隐私泄露风险。与此同时，黄桂新等（2023）强调，即使基于多准则决策框架的智能化平台能够在一定程度上提升决策透明度，但其量化标准的选择和权重分配可能因地域差异而产生偏差，导致潜在的伦理争议。这些挑战提示我们在设计和部署多模态大模型时，需要综合考虑技术可行性与社会接受度之间的平衡。​

最后，全球化背景下的多模态大模型推广还面临着文化适配的难题。正如 Yanis Labrak 等人（2024）[14] 所展示的，即便模型在英语环境中表现出色，但在其他语言中的迁移效果却受到语言结构和医学术语差异的影响。这种文化和技术的双重壁垒要求研究者在模型开发阶段就充分考虑本地化需求，从而提高其在不同语境中的适用性。

3. 多模态医疗健康数据的知识组织

多模态医疗健康数据的知识组织是构建高效医疗模型和提升决策质量的关键环节。随着人工智能技术的发展，如何整合不同形态的医疗数据（如文本、图像和结构化记录）成为研究热点。目前的研究主要围绕多模态数据融合机制、知识增强方法及其在实际医疗场景中的应用展开[11]。同时，也有研究探讨了大型语言模型在处理多模态数据时的表现，尤其是在临床决策支持和急救分诊中的潜力[9][10]。

3.1 多模态数据融合与知识增强

多模态数据融合技术为医疗健康领域提供了全新的解决方案。张殿元和余传明（2024）[13] 提出了一种结合知识增强与多模态融合的医疗视觉问答模型。该模型通过引入外部知识图谱，显著提升了对医学图像和临床问题之间关系的理解能力。研究表明，知识图谱的嵌入不仅增强了文本语义表示，还通过多模态融合层有效捕捉了文本与图像之间的交互信息。例如，在对肺部 X 光影像进行问答时，模型能够结合知识图谱中关于肺部疾病的知识，更准确地回答问题。这种知识增强的方法为复杂医疗问题的解决提供了理论支撑，同时也揭示了跨模态数据整合在提升模型预测性能中的重要作用。​

然而，当前研究仍存在一定局限性。例如，所提出的模型仅在单一语言情境下进行了验证，缺乏多语言数据集的支持。这一不足提示，未来研究需要进一步探索多语言环境下的多模态数据融合策略，以适应全球医疗健康领域的多样化需求。

3.2 大型语言模型在多模态数据处理中的表现

大型语言模型（LLMs）在多模态医疗健康数据处理中的应用逐渐受到关注。Brügge 等（2024）[15] 的研究表明，AI 模拟的病史对话能够显著提高医学生的临床决策能力，特别是在结合结构化反馈的情况下。这种基于语言模型的训练形式不仅为医学生提供了低成本的学习途径，还为其应对真实医疗场景中的复杂问题奠定了基础。例如，医学生可以通过与 AI 模拟的患者进行对话，练习询问病史和诊断疾病的能力。此外，Masanneck 等（2024）[16] 比较了大型语言模型、ChatGPT 以及未受训医生在急诊分诊中的表现。结果发现，尽管现有模型的分诊能力尚未达到专业水准，但其最佳表现已与未受训医生相当。这表明，大型语言模型在处理多模态数据时具有潜在优势，但仍需进一步优化其算法设计和知识整合能力。​

值得注意的是，这些研究揭示了大型语言模型在医疗健康领域的广阔前景，但也暴露了其在专业知识深度和实时适应性方面的不足。例如，模型的表现受限于训练数据的质量和覆盖范围，而医疗领域的数据往往具有高度的异质性和复杂性。因此，如何提升模型在多模态数据处理中的鲁棒性和泛化能力，仍是亟待解决的问题。

4. 基于多模态的医疗模型构建与应用

基于多模态的医疗模型是当前人工智能在医疗领域的重要研究方向，其核心在于整合不同类型的数据源（如文本、图像、信号等），以实现更精准的诊断、决策支持和健康管理。本章节将围绕智能决策平台模型构建和视觉问答模型两个关键方面展开探讨，结合现有研究揭示多模态技术在医疗领域的潜力与挑战[[1](" \l "anchor_1)][13]。

4.1 智能决策平台模型构建

智能决策平台的构建旨在利用多准则分析方法，结合医疗大数据的特点，为复杂医疗场景提供科学决策支持。近年来，随着医疗健康数据类型的多样化和规模的快速增长，传统单一模态的决策工具难以满足临床需求。韩普等（2023）[11] 提出了一种面向多模态医疗健康数据的知识组织模式，通过实体对齐、关系抽取和知识融合，构建了具有语义关联的多模态知识图谱。这一模式为智能决策平台提供了坚实的数据基础，同时强调了多模态数据的有序化处理和知识表示的重要性。例如，在制定治疗方案时，智能决策平台可以基于多模态知识图谱，综合考虑患者的病情、病史、基因数据等多方面信息，提供更科学的决策建议。然而，现有研究中关于如何在动态医疗环境中实现高效数据更新和实时决策支持仍缺乏系统性探讨。​

此外，王玥等（2024）[1] 指出，智能决策平台的设计需要兼顾伦理治理与技术可行性，特别是在数据隐私保护和模型透明性方面。例如，如何确保多模态数据来源的合法性和可信度，以及如何评估模型输出结果的置信度，仍然是亟待解决的问题。尽管已有研究表明知识图谱可以有效提升数据解释能力，但其在实际应用中的可扩展性和鲁棒性仍需进一步验证。

4.2 视觉问答模型

视觉问答模型作为多模态技术的典型应用之一，在医疗影像解读和临床问题解答中展现出巨大潜力。张殿元和余传明（2024）[13] 提出了一种基于知识增强与多模态融合的医疗视觉问答模型，通过引入外部知识图谱和置信度检测机制，显著提升了模型在开放域问答任务中的表现。该模型由多个功能模块组成，包括文本知识增强层、图像嵌入层、多模态融合层、置信度检测层和预测层，能够有效捕捉文本与图像之间的交互关系，并评估输入数据的可靠性。​

实验结果显示，该模型在 VQA - RAD 和 PathVQA 数据集上的开放域问答准确率分别达到了 59.3% 和 16.2%，证明了其在复杂医疗场景中的适用性。然而，作者也指出，当前模型仅考虑了单一语言情境，未来需要在多语言数据集上进行进一步验证。此外，Smeaton（2024）[3] 提出了一个值得思考的问题：多模态大模型是否会对传统多媒体分析方法造成冲击？在医疗视觉问答领域，这一问题尤为重要，因为经典方法往往依赖特定领域的专家知识，而新兴的多模态模型则可能忽略这些背景。

4.3 多模态模型的伦理与治理考量

除了技术和方法论层面的探讨，多模态医疗模型的应用还需关注伦理与治理问题。王玥等（2024）[[1](" \l "anchor_1)]强调，世界卫生组织发布的指南明确了六项原则，用于指导多模态大模型的开发与使用，包括公平性、透明性、责任归属等。这些原则不仅为模型设计者提供了参考框架，也为政策制定者和医疗机构实施监督提供了依据。例如，在多模态视觉问答模型中，置信度检测机制的引入正是为了提升模型的可信度，这与指南中对算法透明性的要求高度契合。

另一方面，尽管多模态模型在提升医疗服务质量和效率方面表现出色，但其潜在风险也不容忽视。例如，单一语言情境的局限性可能导致模型在全球范围内的应用受到限制，而数据偏差问题则可能加剧医疗资源分配的不平等。因此，未来的多模态医疗模型不仅需要技术上的突破，还需要跨学科协作，共同应对伦理与治理方面的挑战。

5. 医疗领域大语言模型研究

医疗领域的大语言模型（LLMs）正在迅速成为人工智能驱动医疗服务的重要组成部分。这些模型通过自然语言处理技术，能够从海量医疗数据中提取知识、生成诊断建议以及优化临床决策流程。本章节将围绕开源预训练模型和微调策略展开讨论，评估其在医疗领域的应用潜力与挑战[[6](" \l "anchor_6)][12]。

5.1 开源预训练模型

近年来，开源预训练模型在医疗领域的研究中展现出巨大潜力。例如，Yanis Labrak 等人（2024）[14] 提出了 BioMistral 系列模型，这是一组专为医疗领域设计的开源预训练大语言模型。该研究通过十项医学问答任务对 BioMistral 进行了全面评估，并将其基准测试翻译为七种语言，首次实现了大规模多语言医疗模型的性能评估。结果表明，BioMistral 在多种语言环境下均表现出了较高的准确性和鲁棒性，为跨语言医疗信息处理提供了坚实基础。在非洲撒哈拉以南地区，当地医疗人员借助 BioMistral 模型，能够快速获取国际前沿的疟疾防治知识，极大提升了疾病防控能力。​

此外，L. Masanneck 等（2024）[16] 的研究揭示了开源模型在急诊医学中的表现。他们的研究表明，尽管现有的大型语言模型尚未完全达到专业医生的水平，但最佳模型的表现已接近未经培训的急诊医生。在模拟的突发公共卫生事件场景中，开源模型可根据患者症状描述，初步判断疾病类型并给出分诊建议，在一定程度上缓解了急诊资源紧张时的人力压力。然而，开源模型的性能仍然受到数据质量和模型架构的限制，尤其是在复杂医疗情境下，如罕见病诊断、多器官功能衰竭等病例的推理能力仍需大幅提升。

5.2 微调策略评估

微调策略的选择对于提升医疗大语言模型的性能至关重要。Cl'ement Christophe等人（2024）[12]对两种主流微调方法——全参数微调和参数高效微调——进行了系统比较。研究结果表明，参数高效微调在资源消耗和模型性能之间取得了较好的平衡，尤其适用于医疗领域中小规模标注数据集的情况。相比之下，全参数微调虽然能够进一步提高模型性能，但其高昂的计算成本和对大规模标注数据的依赖使其难以普及。

此外，Emilia Brügge等（2024）[15]通过随机对照试验探讨了结合AI模拟和结构化反馈的微调方法对学生临床决策能力的影响。研究结果显示，基于AI的历史对话模拟和实时反馈显著提升了医学生的临床决策技能，并为传统教学方法提供了一种低成本补充方案。这种创新性的微调方式不仅增强了模型的实际应用场景适应性，也为未来医疗教育技术的发展提供了新思路。

5.3 多语言能力的扩展

随着全球化进程的加速，多语言能力逐渐成为医疗大语言模型不可或缺的功能之一。Yanis Labrak等人（2024）[14]的研究开创性地探索了这一方向，通过对BioMistral进行多语言自动翻译与评估，验证了其在不同语言环境下的适用性。这一研究成果不仅填补了多语言医疗模型研究的空白，还为跨国医疗服务协作提供了技术支持。然而，值得注意的是，多语言扩展过程中的文化适应性问题仍然需要进一步研究，特别是在涉及敏感医疗信息时，如何确保翻译内容的精准性和语义一致性仍然是一个挑战。

6. 多模态大模型对医疗决策的影响

多模态大模型正在重新定义医疗决策的边界，其在临床培训和急诊分诊中的应用展现了巨大的潜力。通过模拟患者互动与结构化反馈机制，这些模型显著提升了医学生的临床决策能力；同时，其在急诊分诊中的表现也引发了关于专业性与效率之间平衡的深刻讨论。本章节将深入探讨多模态大模型在医疗决策领域的具体影响。

6.1 对医学生临床决策的影响

多模态大模型通过模拟真实患者互动与提供结构化反馈，为医学生的临床决策训练提供了创新路径。Brügge等（2024）[15]的研究表明，基于人工智能的患者模拟能够有效支持临床决策训练（Clinical Decision Making, CDM），特别是当这种模拟与结构化反馈相结合时。他们的随机对照试验进一步证实，这种训练方式不仅成本效益高，还能帮助学生更好地适应实际医疗场景中的复杂对话情境。结合曹建峰和徐艳玲（2024）[17]的观点，多模态模型通过整合文本、语音和图像数据，可以更全面地再现患者病例，从而提升训练的真实性与实用性。然而，这种方法也面临挑战：例如，算法偏见可能导致某些特定类型的患者案例被过度简化或忽视，进而影响决策训练的全面性。此外，尽管现有研究表明这一方法具有较高的教育价值，但其长期效果以及在不同文化背景下的适用性仍需进一步验证。

6.2 在急诊分诊中的表现

急诊分诊作为医疗决策中的关键环节，对效率和准确性提出了极高要求。Masanneck等（2024）[16]的研究对比了多种大语言模型、ChatGPT以及未接受专业训练医生在急诊分诊中的表现，发现尽管最先进的大语言模型尚未达到专业评估者的水平，但其最佳模型的表现已可与未经训练的急诊科医生相媲美。这表明，多模态大模型在一定程度上具备支持紧急医疗决策的潜力。然而，黄桂新等（2023）的研究则强调，仅仅依赖语言模型可能不足以应对复杂的急诊场景[[7](" \l "anchor_7)]。他们提出的基于多准则决策分析（MCDA）的智能化决策平台框架，通过量化决策目标与标准，能够更好地处理急诊分诊中涉及的多维度信息，包括患者的生理指标、病史数据以及实时监测结果。因此，未来的急诊分诊系统可能需要结合多模态大模型与多准则决策方法，以实现更高的精准性和可靠性。需要注意的是，当前研究仍存在局限性，例如样本多样性不足以及缺乏针对特定疾病类型的数据优化，这可能限制模型在高度专业化场景中的表现。

6.3 医疗决策中的伦理与治理问题

随着多模态大模型在医疗决策中的广泛应用，其带来的伦理与治理问题也逐渐显现。曹建峰和徐艳玲（2024）[17]指出，虚假信息传播、情感操纵以及算法偏见等问题亟待解决，这些问题可能严重影响医疗决策的公平性与透明度。尤其是在急诊分诊等高风险场景中，错误的模型输出可能导致严重的医疗后果。因此，研究者提出了一系列治理建议，包括落实全生命周期安全措施、践行伦理嵌入设计的理念，以及明确产品责任规则的适用范围。这些措施不仅有助于增强模型的可信度，还为未来医疗决策系统的可持续发展提供了方向。与此同时，如何在全球范围内协调不同地区的技术标准和伦理规范，也成为一项重要课题。

7. 结论

多模态大模型（MLLMs）在医疗领域的应用探索正逐步揭示其变革性潜力，为应对当代医疗挑战提供了创新解决方案。本文通过系统梳理现有研究，深入探讨了多模态大模型在医疗数据处理、智能决策支持、临床培训以及急诊分诊中的应用，并识别了技术优势与潜在挑战。以下部分将从主要研究发现、现有研究空白和未来研究方向三个方面进行总结。

7.1 主要研究发现

多模态大模型的核心价值在于其整合跨模态医疗数据的能力，充分体现了技术进步对医疗实践的赋能作用。首先，多模态融合技术显著提升了复杂医疗场景中的数据分析能力。例如，韩普等（2023）[11]提出的知识组织模式和张殿元与余传明（2024）[13]开发的视觉问答模型，均展示了通过知识图谱嵌入和多模态交互实现跨模态信息整合的优势。这些研究成果不仅提高了模型对医学图像和临床问题的理解能力，还为精准医疗提供了理论基础。其次，开源预训练模型及其微调策略的优化进一步推动了多模态大模型的实际落地。Yanis Labrak等人（2024）[14]提出的BioMistral系列模型表明，多语言能力的引入能够突破语言障碍，促进全球范围内的医疗协作。此外，Cl'ement Christophe等人（2024）[12]的研究则证明了参数高效微调策略在降低资源消耗的同时保持模型性能的有效性。

与此同时，多模态大模型在临床培训中的应用展示了其在提升医学生决策能力方面的潜力。Brügge等（2024）[15]的研究表明，结合AI模拟和结构化反馈的训练方式不仅能增强学习效果，还为传统教学方法提供了低成本补充。另一方面，在急诊分诊领域，尽管大型语言模型的表现尚未完全达到专业医生水平，但其最佳表现已接近未经培训的急诊科医生[[10](" \l "anchor_10)]。这种进展为进一步优化紧急医疗决策系统奠定了基础。

7.2 现有研究空白

尽管多模态大模型在医疗领域取得了显著进展，但当前研究仍存在一系列局限性和未解决的问题。首先，技术层面的整合难度仍然较高。例如，韩普等（2023）[11]指出，高质量的预处理和实体对齐对于专业知识密集且格式复杂的医疗数据而言具有较高门槛，这限制了多模态数据融合的广泛应用。其次，多语言环境下的适应性问题尚未完全解决。尽管BioMistral模型在多语言任务中表现出色[[6](" \l "anchor_6)]，但文化适配和语义一致性问题仍需进一步研究，特别是在涉及敏感医疗信息时。

此外，伦理与治理问题构成了另一重要挑战。黄桂新等（2023）[[7](" \l "anchor_7)]强调，基于多准则决策分析的智能化平台虽然能够在一定程度上提升透明度，但量化标准的选择可能因地域差异引发争议。同时，曹建峰和徐艳玲（2024）[[2](" \l "anchor_2)]的研究揭示了算法偏见、虚假信息传播及情感操纵等风险，对医疗决策的公平性和可靠性提出了严峻考验。最后，样本多样性和数据覆盖范围的不足也限制了模型在特定疾病类型或高复杂度场景中的表现[16]。

7.3 未来研究方向建议

针对上述研究空白，未来研究应聚焦以下几个方向以推动多模态大模型在医疗领域的全面发展。首先，需要加强多模态数据融合技术的研究，特别是提高其实时性和鲁棒性。未来工作可以探索自动化的实体对齐和关系抽取方法，以降低技术门槛并提高跨模态数据整合效率[[4](" \l "anchor_4)]。其次，多语言能力的扩展应当更加注重文化适配问题，建议开展大规模的跨语言验证实验，确保翻译内容的精准性和语义一致性[14]。

其次，伦理与治理框架的完善至关重要。未来研究应致力于构建全生命周期的安全措施，将伦理嵌入设计作为核心原则之一[[2](" \l "anchor_2)]。同时，跨学科合作能够帮助制定统一的技术标准和伦理规范，从而在全球范围内协调不同地区的治理政策。再次，针对特定疾病类型和高复杂度场景的数据优化也是不可或缺的方向。通过扩充高质量标注数据集，可以进一步提升模型在专业化场景中的表现[12][16]。

最后，研究者应关注多模态大模型与其他先进技术的协同应用。例如，将多模态大模型与多准则决策分析方法相结合，有望打造更高效的急诊分诊系统[7]。此外，与物联网和实时监测技术的集成也可为动态医疗决策提供更全面的支持。

综上所述，多模态大模型在医疗领域的应用前景广阔，但仍需克服技术、伦理和治理等方面的多重挑战。通过持续的创新研究，这一技术有望在未来实现更大规模的实际应用，为全球医疗健康事业的发展注入新的活力。

参考文献

1. [] 王玥, 宋雅鑫, 王艺霏, 于莲, & 王晶. (2024). 卫生领域人工智能的伦理与治理:多模态大模型指南. 中国医学伦理学. [↑](#endnote-ref-0)
2. [] Zhou, J., He, X., Sun, L., et al. (2024). Pre-trained Multimodal Large Language Model Enhances Dermatological Diagnosis using SkinGPT-4. Nature Communications. [↑](#endnote-ref-1)
3. [] Smeaton, A. (2024). The LLM Wrecking Ball: Are We About to Lose Decades of Work in Multimedia because of MM-LLMs? Proceedings of the 2024 International Conference on Multimedia Retrieval.

   1. Li, T.-X., Su, Y.-C., Li, W., et al. (2024). GMAI-VL & GMAI-VL-5.5M: A Large Vision-Language Model for General Medical AI. arXiv.
   2. d Shi, Y., Zhu, X., Hu, Y.et al. (2024). Med-2E3: A 2D-Enhanced 3D Medical Multimodal Large Language Model. arXiv.
   3. Sun, K., Xue, S., Sun, F., et al. (2024). Medical Multimodal Foundation Models: Applications and Challenges. arXiv.
   4. Shen, S., Qi, W., Liu, X., et al. (2025). From Virtual to Reality: Innovative Practices of Digital Twins in Tumor Therapy. Journal of Translational Medicine.
   5. Han, T., Adams, L.C., Bressem, K., et al. (2024). Comparative Analysis of Multimodal Large Language Model Performance on Clinical Vignette Questions. JAMA.
   6. Zhou, J., He, X., Sun, L., et al. (2024). Pre-trained Multimodal Large Language Model Enhances Dermatological Diagnosis using SkinGPT-4. Nature Communications.
   7. Hoque, M., Hasan, Md. R., Emon, Md. I. S., et al. (2024). Medical Image Interpretation with Large Multimodal Models. IEEE Journal of Biomedical and Health Informatics.
   8. 韩普，叶东宇，陈文褀，等. (2023). 面向多模态医疗健康数据的知识组织模式研究。现代情报.
   9. Christophe, C., Kanithi, P., Munjal, P., et al. (2024). Med42 - Evaluating Fine-Tuning Strategies for Medical LLMs. arXiv.
   10. 张殿元，余传明. (2024). 基于知识增强与多模态融合的医疗视觉问答模型。数据分析与知识发现.
   11. Labrak, Y., Bazoge, A., Morin, E., et al. (2024). BioMistral: A Collection of Open-Source Pretrained Large Language Models. arXiv.
   12. Brügge, E., Ricchizzi, S., Arenbeck, M., et al. (2024). Large language models improve clinical decision making. BMC Medical Education.
   13. Masanneck, L., Schmidt, L., Seifert, A., et al. (2024). Triage Performance Across Large Language Models. Journal of Medical Internet Research.
   14. 曹晓明，张永和，潘萌，等. (2021). 人工智能视域下的学习参与度识别方法研究。远程教育杂志.
   15. 李锡荣. (2021). 多模态深度学习及其在眼科人工智能的应用展望。协和医学杂志.
   16. Liu, F., Zhu, T., Wu, X., et al. (2023). A Medical Multimodal Large Language Model for Future Pandemics. NPJ Digital Medicine.
   17. 龚瑾. (2023). 人工智能赋能的多模态学习状态研究与分析。信息系统工程.

   [↑](#endnote-ref-2)