

医学专家大模型调研报告

一、引言

1.1 医学专家大模型：定义、范畴与核心议题

医学专家大模型（Medical Expert Large Models）是指在人工智能（AI）领域中，那些具备高级别医学知识理解、复杂医学逻辑推理与精准临床决策辅助能力的先进AI系统。这些模型不仅局限于处理特定任务，如医学影像分析的深度学习模型，也涵盖了能够进行自然语言交互、海量医学知识检索与内容生成的医学大型语言模型（LLM）。“专家”一词凸显了其在特定医学场景下，能力接近甚至在某些方面超越人类医学专家的潜力。

本报告的范畴将重点探讨两条核心技术路径下的医学专家大模型：

1. **基于深度学习的医学影像分析模型**：这类模型，特别是利用卷积神经网络（CNN）、U-Net 及其变体、以及视觉 Transformer（ViT）等架构，在医学图像的病灶检测、分割、分类、良恶性判断以及预后评估等任务中取得了显著成就。在某些特定诊断任务中，例如特定类型癌症的影像学早期筛查与精准分级，其表现已展现出专家级水准¹。
2. **面向医疗领域的自然语言处理与大型语言模型**：这类模型专注于理解和处理复杂的医学文本信息，如电子病历、医学文献、临床指南和医患对话。通过在海量医学文本上进行预训练和微调，医学 LLM（如 Med-PaLM 系列、GatorTron 等）在临床决策支持、医学知识问答、病历自动摘要与生成、药物研发辅助以及改善医患沟通等方面显示出巨大应用潜力³。

“医学专家大模型”这一概念本身也经历了一个演化和融合的过程。最初，医学AI的“专家”能力更多体现在如计算机辅助诊断（CAD）这类由规则或早期机器学习算法驱动的、针对特定任务的辅助系统上⁶。这些系统旨在提高医生诊断的敏感性或特异性。随着深度学习技术的兴起，尤其是在医学影像分析领域，AI模型开始在特定任务（如癌症筛查、病灶识别）上达到甚至超越人类专家的平均水平，这标志着第一波真正意义上的“专家模型”的出现¹。这些模型通常是针对单一模态（主要是影像）和特定病种的。

近期，大型语言模型（LLMs）的突破为“医学专家”的内涵注入了新的维度。通用 LLMs 在标准化医学知识问答（如美国执业医师资格考试 USMLE）上展现出惊人能力⁴，这催生了专门针对医学领域优化的 LLMs，例如谷歌的 Med-PaLM 系列和佛罗里达大学的 GatorTron⁴。这些模型主要处理的是文本信息，其能力更多体现在模拟医生的认知、推理、知识整合和沟通能力上。

用户对“医学专家大模型”的关注，实际上暗示了对这两种核心能力——即强大的感知智能（如影像分析）和高级的认知智能（如语言理解与医学推理）——的综合期望。学术界和产业界也已开始积极探索多模态融合技术，例如结合影像学特征和临床文本数据进行更精准的诊断或预后预测²。因此，“医学专家大模型”的概念正从指代在单一任务上表现卓越的 AI，逐渐演变为一种能够整合多源异构医疗信息、具备更全面认知能力和复杂决策辅助能力的综合性人工智能系统。这种概念的演化深刻反映了 AI 技术发展的内在逻辑——从感知智能向认知智能的深化，以及对更复杂、更接近人类医生综合诊疗能力的持续追求，预示着未来的医疗 AI 将更加强大和全面。

本报告的核心议题将围绕以下问题展开：这些医学专家大模型是如何学习并尝试模拟，乃至在特定方面超越人类医学专家的能力的？它们在发展过程中面临哪些关键的技术瓶颈与伦理挑战？其商业化落地的现状与前景如何？以及，它们将如何深刻重塑未来的医疗健康服务体系与医学研究范式？

1.2 本报告调研范围与结构概述

本报告旨在对医学专家大模型进行一次全面而深入的调研。调研内容将覆盖其技术演进脉络、核心应用价值与场景、当前学术研究的前沿动态与尚待突破的瓶颈、发展历程中的关键里程碑事件，以及备受关注的产业化实践案例。通过多维度的剖析，力求为读者呈现一幅关于医学专家大模型发展现状与未来趋势的清晰图景。

报告结构安排如下：

- **第二章** 将阐述医学专家大模型诞生的背景，分析驱动其发展的核心因素，并详细介绍

绍其关键能力及在医疗健康领域的核心应用价值。

- **第三章** 将聚焦于学术研究层面，探讨当前医学专家大模型研究的热点方向，并深入分析其面临的主要技术瓶颈和深层次挑战。
- **第四章** 将回顾医学专家大模型的发展历程，梳理从早期计算机辅助诊断到现代深度学习和大型语言模型的关键里程碑。
- **第五章** 将转向产业化视角，剖析医学影像 AI 和医学大型语言模型的商业化实践，并列举代表性企业及其解决方案。
- **第六章** 将对全文进行总结，凝练核心研究发现，并对医学专家大模型的未来发展趋势、潜在突破口及相关战略提出展望与建议。
- **第七章** 为参考文献列表。

二、医学专家大模型的演进背景及核心应用价值

2.1 驱动因素与发展脉络

医学专家大模型的兴起与快速发展，是医疗领域固有痛点与人工智能技术突破性进展双重驱动的结果。这种“需求拉动”与“技术推动”形成的双轮效应，共同塑造了当前医学 AI 的蓬勃态势。

医疗领域对智能化解决方案的迫切需求 (需求拉动)

医疗健康领域面临着一系列严峻挑战，这些挑战构成了对智能化解决方案的强烈需求：

- **医疗数据爆炸式增长**：随着影像技术、基因测序、电子病历系统的普及，医疗数据量正以前所未有的速度激增。传统的人工处理和分析方式已难以有效应对海量、高维度、多模态的医疗数据，从中提取有价值的临床洞见变得日益困难¹⁰。
- **提升诊疗效率与质量的压力**：全球范围内，医疗系统普遍面临提高诊断准确率、降低医疗差错、缩短患者等待时间的压力。误诊和漏诊不仅可能对患者造成严重后果，也增加了医疗系统的负担²。
- **医疗资源分布不均与可及性问题**：优质医疗资源（尤其是经验丰富的专家）往往集中

在大型医疗机构和发达地区，基层和欠发达地区的医疗服务能力相对薄弱。AI 有潜力通过赋能基层医生、提供远程辅助等方式，缓解资源不均问题¹²。

- **日益增长的医疗服务需求与成本控制挑战**：人口老龄化、慢性病发病率上升等因素导致医疗服务需求持续增长，而医疗费用的不断攀升也给各国医保体系带来巨大压力。AI 有望通过提高效率、优化资源配置来帮助控制成本¹²。
- **临床医生工作负荷沉重与职业倦怠**：临床医生，特别是放射科、病理科等科室的医生，面临着繁重的工作量。例如，有研究指出，医生平均花费在临床文档记录上的时间几乎是直接提供患者护理时间的两倍¹⁴。这种行政负担不仅占用了宝贵的诊疗时间，也是导致医生职业倦怠的重要原因¹⁰。AI 驱动的自动化工具，如自动生成病历摘要、辅助阅片等，被寄予厚望以减轻医生负担，提升工作满意度¹⁶。

人工智能技术的突破性进展 (技术推动)

与此同时，人工智能技术的飞速发展为解决上述医疗领域的痛点提供了强大的技术支撑：

- **深度学习算法的成熟与广泛应用**：自 21 世纪 10 年代以来，以卷积神经网络 (CNN) 为代表的深度学习算法在图像识别等领域取得巨大成功。随后，Transformer 架构的提出及其在自然语言处理 (NLP) 领域的革命性影响，为大型语言模型 (LLM) 的诞生奠定了基础¹。这些先进算法能够从大规模数据中自动学习复杂模式和特征，是医学专家大模型能力的核心。
- **计算能力的飞跃与大规模数据集的可及性增强**：图形处理器 (GPU) 等并行计算硬件的快速发展，极大地提升了训练复杂深度学习模型所需的计算能力¹⁵。同时，随着医疗信息化的推进和一些公开医学数据集（如 TCGA、MIMIC 等）的建立，研究人员能够获取到更大规模的数据用于模型训练，尽管高质量标注医学数据的获取仍是一大挑战⁶。
- **大型语言模型的出现及其卓越能力**：基于 Transformer 架构的大型语言模型，通过在海量文本数据上进行预训练，展现出在理解自然语言、生成连贯文本、进行知识推理等方面的强大能力³。这为开发能够理解医学文献、处理临床笔记、与医患进行自然交互的 AI 系统开辟了全新路径。

需求驱动与技术推动的结合，形成了一个正向循环。医疗领域的迫切需求为 AI 技术的研发设定了明确的应用目标和检验其价值的真实场景。例如，医生对减少繁琐文书工作的强烈需求¹⁴，直接拉动了 AI 辅助临床文档记录技术（如 AI Scribe）的发展²²。反过来，AI 技术的每一次突破，如 Transformer 架构的出现和 LLM 能力的增强³，都为解决医疗难题开辟了新的可能性，甚至催生出以往难以想象的应用场景，例如个性化的大规模患者教育和精准的药物研发辅助。这种双轮驱动效应确保了医学专家大模型发展的持续动力和明确方向，预示着技术突破将优先应用于解决最迫切的临床需求，而未被满足的临床需求也将持续激励技术的进一步创新。

2.2 医学专家大模型的关键能力

医学专家大模型的核心能力主要体现在两大技术方向：基于深度学习的医学影像分析和面向医疗领域的自然语言处理与大型语言模型。

基于深度学习的医学影像分析

深度学习模型，特别是卷积神经网络（CNNs）、U-Net 及其变体，以及近年来兴起的视觉 Transformer（ViT），在医学影像分析领域展现出强大的能力，能够辅助甚至在某些方面超越人类专家：

- **病灶检测、分割、分类与量化评估**：AI 模型能够自动识别医学影像（如 X 光片、CT、MRI、超声图像、数字病理切片等）中的异常区域，如肿瘤、结节、出血、骨折等。它们不仅能检测病灶的存在，还能精确地分割出病灶的轮廓、计算其大小、体积、密度等量化指标，并对其进行良恶性分类或疾病分级¹。例如，在癌症筛查中，AI 能够辅助医生进行早期检测，对肿瘤进行精准分级，从而提高诊断的准确性和一致性¹。
- **支持多种影像模态**：AI 技术的应用已覆盖几乎所有主流医学影像模态，包括二维的 X 射线、乳腺 X 线摄影，以及三维的计算机断层扫描（CT）、磁共振成像（MRI）、正电子发射断层扫描（PET），还有动态的超声心动图、内窥镜视频，以及微观层面的数字病理图像等²。

- 代表性模型架构：

- **CNNs**：作为图像识别的基石，CNN 通过其卷积层、池化层和全连接层有效提取图像的局部和全局特征，广泛应用于各类医学影像的分类和检测任务²³。
- **U-Net**：这是一种专门为生物医学图像分割设计的编码器-解码器架构，其 U 形结构和跳跃连接能够有效地结合深层语义信息和浅层细节信息，在像素级分割任务中表现优异²⁷。
- **Vision Transformers (ViT)**：借鉴了 NLP 领域 Transformer 模型的成功，ViT 将图像分割成小块（patches）并将其作为序列输入，通过自注意力机制捕捉图像中的长程依赖关系，在一些医学影像分析任务中展现出超越 CNN 的潜力²。

面向医疗领域的自然语言处理与大型语言模型

大型语言模型（LLMs）及其在自然语言处理（NLP）方面的能力，为处理和利用医疗领域中海量的文本信息提供了前所未有的机遇：

- **理解和处理非结构化医疗文本**：医疗信息中有高达 80% 是以非结构化文本形式存在的，如医生的临床笔记、出院小结、病理报告、影像报告、医学文献以及医患之间的对话记录等³。医学 LLMs 经过专门训练，能够理解这些文本中复杂的医学术语、缩写、上下文关系以及潜在的语义信息³。
- **生成类人回应，支持医患沟通、虚拟助手、患者教育**：LLMs 能够生成流畅、自然且与上下文相关的文本，这使其能够被用于开发智能聊天机器人、虚拟健康助手等应用。这些工具可以 7x24 小时回答患者的常见问题、提供健康咨询、解释复杂的医学概念、进行用药提醒，从而改善患者体验、提升患者的健康素养和自我管理能力³。
- **医学知识检索、信息提取、临床笔记摘要与生成**：LLMs 可以快速从海量的医学文献、临床指南和数据库中检索特定信息，辅助医生和研究人员获取最新知识。它们还能从冗长的临床笔记中自动提取关键信息（如诊断、症状、用药、检查结果），生成结构化的摘要，或辅助医生快速撰写病历文书，从而减轻行政负担³。
- **辅助药物研发与临床试验**：LLMs 能够分析海量科研文献和生物医学数据，帮助识别新的药物靶点、预测候选化合物的活性、毒性和药代动力学特性，从而加速早期药物

筛选过程³²。此外，LLMs 还可以通过分析患者病历和临床试验的入排标准，智能匹配合适的受试者，提高临床试验的招募效率和成功率³²。

- **检索增强生成（RAG）技术提升准确性与可靠性**：为了克服 LLMs 可能产生“幻觉”（即生成不准确或虚假信息）的问题，并确保其输出基于可靠的医学证据，检索增强生成（RAG）技术被广泛研究和应用。RAG 通过在 LLM 生成答案之前，先从可信的外部知识库（如医学教科书、临床指南、最新研究论文）中检索相关信息，并将这些信息作为上下文提供给 LLM，从而显著提高其回答的准确性、事实性和可追溯性²¹。

2.3 核心应用场景与价值体现

医学专家大模型凭借其强大的影像分析和自然语言处理能力，正在医疗健康的多个核心场景中展现出巨大的应用价值，其影响正从单一的技术点突破，演变为对整个医疗流程的系统性赋能。

提升诊断准确性与效率

这是医学 AI 最早也是最受关注的应用领域。

- **癌症筛查与早期诊断**：AI 在多种癌症的影像诊断中显示出卓越性能。例如，在乳腺癌筛查中，AI 系统辅助放射科医生判读乳腺 X 线片，不仅能够维持甚至提高癌症检出率，还能显著减少医生的阅片工作量，一项研究表明工作量可减少近一半²⁴。AI 在识别微小钙化灶、致密性乳腺中的早期病变等方面具有优势³⁶。在肺癌筛查方面，推想医疗的 InferRead CT Lung 等 AI 产品能够自动检测小至 4 毫米的肺结节，并进行密度分析和动态随访比对，辅助早期诊断¹⁰。AI 的应用旨在减少漏诊和误诊，提高早期癌症的发现率，从而改善患者预后¹。
- **病理诊断辅助**：数字病理的兴起为 AI 应用提供了新大陆。Paige.ai 等公司的 AI 产品能够辅助病理医生在显微镜图像中识别癌细胞、进行肿瘤分级分型，如其 Paige Prostate Detect 已获得 FDA 批准用于前列腺癌的辅助诊断³⁸。这有助于提高病理诊断的一致性和效率。
- **减少诊断时间**：AI 模型通常能在数秒或数分钟内完成复杂的影像分析或病历摘要任

务，远快于人工处理速度¹⁰。例如，一个 AI 模型能在几秒到几分钟内完成非小细胞肺癌的分期，而放射科医生可能需要 10-30 分钟，且 AI 准确率可达 88.84%¹⁰。

赋能医学研究与药物发现

AI 正在成为加速医学创新和新药研发的强大引擎。

- **加速文献回顾与知识发现**：医学文献数量呈指数级增长，人类研究者难以实时追踪所有最新进展。LLMs 能够快速阅读、理解和总结海量医学文献，从中提取关键信息、发现潜在关联，为科研人员提供有力的知识支持²⁰。
- **药物靶点识别与化合物筛选**：AI 模型可以通过分析蛋白质结构、基因表达数据、化合物性质等信息，预测新的药物靶点，并从庞大的化合物库中筛选出具有潜力的候选药物分子，优化药物设计流程³。例如，Exscientia 公司利用 AI 设计的药物分子已经成功进入人体临床试验阶段，标志着 AI 在药物发现领域的实际突破³。
- **临床试验优化**：临床试验的患者招募往往耗时且成本高昂。AI 可以通过分析电子病历数据和试验方案，自动筛选符合入排标准的潜在受试者，从而提高招募效率和试验成功率³²。例如，TrialGPT 在将患者与合适的临床试验进行匹配方面，准确率达到 87%，并且能够将筛选速度提高 40%³²。

优化临床工作流程与决策支持

AI 不仅提升单点任务的性能，更重要的是能够深度融入临床工作流程，优化整体效率。

- **自动化临床文档**：临床文档记录是医生主要的行政负担之一。以 Nuance DAX Copilot 和 Abridge 为代表的 AI 医疗“书记员”（AI Scribe）解决方案，利用环境计算和 LLM 技术，能够实时捕捉医患对话，并自动将其转换为结构化的临床笔记或病历摘要，然后无缝集成到电子病历（EHR）系统中¹⁴。这极大地减少了医生在文书工作上花费的时间，使他们能更专注于患者。
- **智能临床决策支持（CDS）**：LLMs 可以作为强大的临床决策辅助工具。例如，Glass Health 平台利用 AI 分析患者摘要，生成鉴别诊断列表和初步的诊疗计划草

案，供医生参考³。纽约大学朗格尼健康中心 (NYU Langone Health) 开发的 NYUTron 模型则能够预测患者在出院后 30 天内再入院的风险，帮助医生提前干预³。

- **集成电子病历 (EHR) 系统**：领先的 EHR 供应商如 Epic Systems 正积极将 AI 功能深度嵌入其平台。这些功能包括智能提取病历信息、自动为患者消息起草回复、辅助医嘱录入、进行临床路径推荐等，旨在使 EHR 系统从单纯的信息记录工具转变为智能化的工作伙伴¹⁶。

推动个性化医疗与患者管理

AI 使医疗服务更加精准化和以患者为中心。

- **基于个体数据的精准治疗推荐**：LLMs 能够整合患者的病史、生活习惯、基因组学信息、影像学特征等多维度数据，结合最新的医学知识和临床指南，为患者提供更加个性化的治疗方案建议和预后预测³。
- **改善患者沟通与教育**：LLMs 驱动的聊天机器人或虚拟健康助手，能够用通俗易懂的语言向患者解释复杂的医学概念、回答他们的疑问、提供用药指导和健康管理建议，从而提升患者的参与度和依从性³。
- **远程患者监护与风险预警**：AI 可以分析来自可穿戴设备、家庭监护设备以及患者自我报告的健康数据，实时监测患者的生理状况和病情变化，对潜在的健康风险（如心律失常、血糖波动、病情恶化）进行早期预警，并提示医护人员或患者采取主动干预措施³。

医学 AI 的价值实现，正经历着从早期聚焦于特定诊断任务的“单点技术突破”（如 CAD 系统提高结节检出率⁶ 或深度学习模型在图像分类上的高准确率¹），向“系统性赋能医疗全流程”的深刻转变。当前，随着模型能力的增强和应用场景的多样化（影像 AI 与 LLM 的协同），AI 的应用已扩展到诊断、治疗、研究、管理等多个环节¹。更重要的是，AI 开始深度融入临床工作流，例如 AI scribe 直接改变医生的文档记录方式²²，EHR 集成 AI 优化信息交互⁴⁵，CDS 辅助核心决策过程⁴³。这表明 AI 不再仅仅是一个孤立的辅助工具，而是逐渐成为医疗流程中不可或缺的有机组成部分。其价值体现也从单纯的技术指标

(如诊断准确率 AUC 值) 扩展到更广泛、更实际的临床和经济效益, 例如显著减少医生倦怠¹⁸、大幅提高工作效率¹⁰、增加患者接诊量和服务能力⁴¹、提升患者满意度和就医体验¹⁸, 乃至改善医疗机构的整体运营效率和投资回报率 (ROI)⁴¹。这种转变标志着 AI 的价值正在从“高效工具”向“赋能平台”演进, 其影响是系统性的, 覆盖从预防、诊断、治疗到患者管理和医学科研的整个医疗价值链。这种系统性赋能是医学 AI 走向成熟和大规模应用的关键, 它不仅要求 AI 技术本身足够强大, 更要求其能够与现有医疗生态系统良好适配和深度融合, 最终带来全局性的效率提升和医疗质量改善。

三、医学专家大模型的学术研究前沿与瓶颈挑战

医学专家大模型的飞速发展离不开学术界的持续探索和创新。然而, 在迈向更广泛、更深入临床应用的过程中, 这些模型也面临着诸多技术瓶颈和深层挑战。

3.1 当前学术研究热点方向

先进模型架构探索

学术界持续致力于研发和优化适用于医疗领域的 AI 模型架构, 以提升其性能、效率和特异性。

- **Transformer 及其变体的主导地位** : Transformer 架构因其强大的序列处理能力和对长程依赖关系的捕捉, 不仅在自然语言处理领域催生了大型语言模型, 也在医学影像分析领域 (如 Vision Transformers - ViT) 展现出巨大潜力²⁹。ViT 通过将图像视为一系列图块 (patches) 并利用自注意力机制, 能够比传统卷积神经网络 (CNNs) 更好地学习图像的全局上下文信息和空间相关性²⁹。
- **U-Net 及其改进型的持续优化** : U-Net 架构及其衍生模型 (如 UNet++) 凭借其在医学图像分割任务中的优异表现, 仍然是研究的热点²⁷。UNet++通过引入嵌套和密集的跳跃连接, 旨在进一步缩小编码器和解码器子网络间特征图的语义差距, 从而提升分割精度²⁸。
- **专为医疗领域设计的 LLMs** : 针对医疗文本的独特性和医学知识的复杂性, 研究者们开发了专门的医学大型语言模型。例如, 谷歌的 Med-PaLM 系列 (Med-PaLM, Med-PaLM 2) 通过在海量医学文献和临床数据上进行微调, 在回答医学资格考试 (如 USMLE) 题目等任务上取得了令人瞩目的成绩, Med-PaLM 2 在 MedQA 数据

集上的准确率达到了 86.5%，相较于第一代 Med-PaLM 提升了超过 19%⁴。佛罗里达大学开发的 GatorTron 系列模型，基于 BERT 架构，参数量最大达到 89 亿，在多项临床自然语言处理（NLP）任务（如自然语言推断 NLI 和医学问答 MQA）上显著优于早期的 BioBERT 和 ClinicalBERT 等模型⁵。

- **定制化模型架构**：此外，针对特定的医疗任务（如罕见病诊断、特定类型影像分析）或数据特点（如小样本、多噪声），研究人员也在不断探索和设计更具针对性的定制化模型架构。

多模态信息融合与处理

真实的临床决策往往依赖于多种来源的信息。因此，如何有效地融合和处理多模态医疗数据，是当前医学 AI 研究的重要前沿。

- **整合异构医疗数据**：研究者致力于开发能够整合来自不同来源的医疗数据模型，例如结合医学影像（CT、MRI 等）、数字病理图像、基因组学数据、蛋白质组学数据、临床化验结果、电子病历中的文本记录、甚至可穿戴设备收集的生理信号等，以期提供更全面、更精准的诊断、预后评估和治疗建议²。
- **特征表示的对齐与融合**：多模态融合的关键挑战在于如何有效地对齐不同模态数据的特征表示，并将其融合到一个统一的框架中进行联合分析和学习。
- **LLM 在多模态处理中的潜力**：大型语言模型在处理和理解文本信息方面的强大能力，使其在多模态融合中也扮演着重要角色。例如，研究者正在探索如何利用 LLMs 结合影像特征和对应的放射学报告文本进行更深层次的分析和理解⁹。新兴的多模态 LLM 在增强医学影像分析方面显示出特别的前景³¹。

模型可解释性、可信度与鲁棒性研究

在医疗这一高风险领域，模型的“黑箱”特性是其临床应用的主要障碍之一。因此，提升模型的可解释性、可信度和鲁棒性至关重要。

- **可解释人工智能（XAI）技术**：XAI 旨在阐明 AI 模型的内部工作机制和决策依据，从而增强临床医生对模型输出结果的理解和信任。在医学影像领域，常用的 XAI 方法包括生成显著性图（如 Grad-CAM、XGradCAM）来高亮显示对模型决策贡献最大的图像区域，或使用 LIME、SHAP 等技术来解释个体预测⁴⁹。例如，XGradCAM 在

有效高亮医学图像中的相关异常区域方面表现出较好的效果⁴⁹。

- **检索增强生成 (RAG)**：对于 LLMs, RAG 技术通过在生成答案前从可信的外部知识库中检索相关信息，并将其作为上下文输入到模型中，从而显著提升了模型输出的事实准确性和可追溯性，有效减少了“幻觉”现象的发生²¹。这对于确保 LLM 在医疗场景下提供可靠信息至关重要。
- **模型鲁棒性与泛化能力**：研究模型在面对训练数据中未见过的新数据、不同采集设备或参数、不同人群特征时的表现稳定性（即泛化能力），以及模型在受到微小扰动或对抗性攻击时的抵抗能力（即鲁棒性）⁴⁹。
- **不确定性量化**：开发使模型能够评估并表达其预测结果置信度的方法，这对于临床医生判断何时可以依赖 AI 的建议至关重要。

数据高效利用与隐私保护技术

医学数据的获取成本高昂且常常涉及隐私问题，因此研究数据高效利用和隐私保护技术具有重要意义。

- **数据高效学习方法**：鉴于高质量、大规模标注医学数据的稀缺性，小样本学习 (Few-shot learning)、迁移学习 (Transfer learning)、自监督学习等技术受到广泛关注。这些技术旨在利用有限的标注数据或无标注数据进行有效学习²。
- **隐私保护计算技术**：联邦学习 (Federated learning) 允许多个医疗机构在不直接共享各自原始患者数据的前提下，协同训练一个共享的 AI 模型，从而在保护数据隐私的同时汇聚更多数据资源²³。此外，差分隐私、同态加密等技术也在医学 AI 领域得到探索。
- **数据增强 (Data augmentation)**：通过对现有医学数据进行旋转、裁剪、添加噪声等变换，或利用生成对抗网络 (GANs) 等技术合成新的训练样本，以扩充训练数据集的规模和多样性，提升模型性能²。

3.2 面临的主要技术瓶颈与深层挑战

尽管医学专家大模型取得了显著进展，但其在走向广泛临床应用的过程中仍面临一系列严峻的技术瓶颈和深层次挑战。这些挑战不仅涉及技术层面，也关乎伦理、法律和社会接受度。

数据获取与治理困境

数据是 AI 的基石，但在医学领域，数据问题尤为突出。

- **高质量标注数据的缺乏**：训练高性能的监督学习模型需要大量经过专家精确标注的医学数据。然而，医学数据的标注过程耗时、费力且成本高昂，需要具备专业知识的医生参与。因此，高质量、大规模、多样化的标注医学数据仍然是制约模型性能和泛化能力的主要因素之一²。
- **数据孤岛与标准化难题**：医疗数据往往分散存储在不同医疗机构的异构信息系统中，形成“数据孤岛”。由于缺乏统一的数据标准、接口和共享机制，跨机构的数据整合与共享极为困难，这限制了构建更大规模、更具代表性训练数据集的可能性¹¹。
- **数据隐私与安全**：患者数据是高度敏感的个人信息，其使用受到严格的法律法规保护（如美国的 HIPAA、欧盟的 GDPR 等）。在数据采集、存储、处理和共享过程中，必须采取严密的隐私保护和安全措施，防止数据泄露和滥用⁵⁰。
- **训练数据偏见**：用于训练 AI 模型的数据如果未能充分代表目标应用人群的多样性（例如，在种族、性别、年龄、社会经济状况等方面存在偏倚），则可能导致模型产生带有偏见的输出，对某些特定人群的诊断或治疗建议不准确甚至有害，从而加剧现有的健康不平等问题⁵⁰。例如，多数大型语言模型主要基于标准英语文本进行训练，这可能导致其在处理非英语语言或特定方言时的性能下降⁵⁰。

模型泛化能力与临床适用性差距

AI 模型在实验室环境下的优异表现，并不总能直接转化为真实临床环境中的可靠应用。

- **过拟合与泛化能力不足**：模型可能过度拟合训练数据集的特性，导致其在训练集上表现良好，但在从未见过的外部独立数据集或真实世界的临床数据上性能显著下降⁴⁹。
- **数据异质性的挑战**：真实临床环境中的数据具有高度异质性，例如来自不同品牌和型号的医疗设备、不同的扫描参数设置、不同疾病阶段的患者、以及不同人群的生理特征差异等。这些因素都可能影响模型的稳定性和泛化能力⁴⁹。
- **从研究到临床的转化鸿沟**：将一个在研究中表现良好的 AI 原型转化为一个在临床实践中安全、有效、可靠且易于集成的产品，需要经历漫长而复杂的过程，包括严格的临床验证、监管审批、以及与现有临床工作流程的整合。

“幻觉”现象与算法偏见问题

这是大型语言模型和某些深度学习模型在医学应用中尤为突出的问题。

- **LLM 的“幻觉”**：大型语言模型有时会生成看似合理但实际上错误、捏造或缺乏事实依据的信息，这种现象被称为“幻觉”⁵⁰。在医疗这种对准确性要求极高的领域，模型的“幻觉”可能导致错误的诊断、不当的治疗建议，甚至对患者造成伤害。
- **算法偏见的持续影响**：如前所述，源于训练数据的偏见会直接传递给 AI 模型，导致其在决策时对特定人群产生系统性的不利影响。识别和消除算法偏见是一个持续的挑战。

临床验证的复杂性与标准化评估体系的缺失

如何科学、有效地评估医学 AI 的临床价值和安全性，是其能否被广泛接受和应用的关键。

- **缺乏临床相关的基准和前瞻性研究**：目前许多 AI 模型的评估仍依赖于回顾性数据集或实验室指标，缺乏在真实临床环境中进行的大规模、多中心、前瞻性的临床试验证据来充分证明其临床效用和安全性¹²。现有研究中，专家意见或回顾性分析占比较大，这限制了研究结果的普适性¹²。
- **评估指标与临床实用性的脱节**：传统的自然语言处理评估指标（如 BLEU、ROUGE）或图像识别指标（如 AUC）可能无法完全反映模型在复杂临床场景中的真实价值和对患者结局的实际影响⁵⁰。
- **标准化评估与报告指南的缺失**：医学 AI 领域需要建立公认的临床验证标准、评估平台和报告指南（例如，针对 LLM 的 TRIPOD-LLM 倡议），以规范研究设计、提高研究质量、增强结果的可比性和可信度¹²。

伦理、法律与社会影响（ELSI）及监管框架的滞后性

AI 技术的快速发展也带来了一系列复杂的伦理、法律和社会问题，而相应的监管框架往往滞后于技术进步。

- **责任归属问题**：当 AI 辅助的决策导致医疗差错或不良事件时，其法律责任应由 AI 开发者、医疗机构、开具医嘱的临床医生，还是 AI 系统本身（如果可能的话）来承担？这一问题目前尚无明确答案²⁶。

- **患者安全与知情同意**：如何确保 AI 工具在临床应用中的安全性，避免对患者造成潜在伤害？在 AI 辅助诊疗过程中，如何充分保障患者的知情同意权，使其了解 AI 在其中扮演的角色、潜在风险和获益？⁵⁰。
- **对医患关系的影响**：过度依赖 AI 或不当使用 AI 工具，可能会削弱医生的临床自主判断能力，或者导致医疗过程的“去人化”，影响医患之间的信任和沟通⁵⁰。
- **监管框架的挑战**：现有的医疗器械监管框架大多是为传统的、静态的医疗设备设计的，难以完全适应快速迭代、具有持续学习能力的自适应算法和大型语言模型²⁶。例如，美国食品药品监督管理局（FDA）目前对于这类能够持续学习和演进的 AI 模型的监管方法仍在积极探索和完善中⁵³。

在所有这些瓶颈和挑战中，“信任”问题可以被视为核心中的核心。无论是数据获取的困难、模型泛化能力的不足、LLM 的“幻觉”现象，还是算法偏见的存在²，其最终结果都可能导致模型输出的不可靠或不可预测性。这种不可靠性直接侵蚀了临床医生和患者对 AI 系统的信任¹⁵。如果医生不信任 AI 提供的建议，那么即使 AI 技术本身再先进，也无法有效地整合到临床工作流程中并发挥其应有的价值。当前学术研究的热点，如可解释人工智能（XAI）⁴⁹ 和检索增强生成（RAG）³⁴，其根本目的之一就是通过提高模型的透明度、可解释性和事实准确性来建立和增强信任。同样，临床验证的复杂性和标准化评估体系的缺乏¹²，也因为无法提供充分的、公认的证据来证明 AI 的有效性和安全性，而阻碍了信任的建立。伦理考量和监管框架的不完善²⁶，例如责任归属的不明确和对潜在隐私泄露的担忧，也会进一步降低公众和专业人士对 AI 的信任度。因此，尽管数据、算法、验证等都是具体的技术或流程层面的瓶颈，但它们最终都汇聚到“信任”这一更为根本的障碍上。解决信任问题，是医学专家大模型得以大规模推广和应用的首要前提。

另一个值得关注的现象是，尽管学术界对多模态信息融合（即整合来自影像、文本、基因组学等多种来源的数据进行综合分析）寄予厚望，认为其能够更全面地模拟人类专家的复杂决策过程，从而带来更精准的医疗服务²，但在实际的产业化和临床应用中，这一宏伟目标与能够产生广泛认可的真实世界证据之间，仍存在显著的鸿沟。当前已获得监管批准并得到较广泛应用的医学 AI 产品，尤其是在影像诊断领域，绝大多数仍然是基于单一模

态数据（例如，仅分析 CT 影像的肺结节检测系统³⁷，或仅分析数字病理切片的癌症诊断辅助工具³⁸）。大型语言模型的应用也主要集中在处理文本数据，如临床笔记摘要、医患沟通等³。构建和验证多模态大模型的复杂性远超单模态模型，这涉及到跨模态数据的精确对齐、异构特征的有效融合、模型解释性难题的加剧，以及最为关键的——收集大规模、高质量、多中心、且包含完整多模态信息的患者队列数据集的巨大挑战。临床验证的复杂性¹²在多模态场景下会进一步放大。如何设计严谨的临床试验来清晰地证明多模态模型相对于成熟的单模态模型或人类专家的增量价值，并将其转化为临床实践指南，是一个漫长且充满不确定性的过程。因此，尽管多模态融合无疑是医学 AI 未来的重要发展方向，但从前沿的学术研究热点到产生坚实的、能够指导临床实践的真实世界证据，并最终实现大规模临床应用，中间仍需跨越巨大的技术和实践障碍。这提醒我们在积极关注和投入多模态等前沿研究的同时，也应清醒认识到其临床转化的长期性和艰巨性。在短期内，持续优化和推广那些已被临床验证有效的单模态 AI 应用，积累更多的真实世界数据和应用经验，可能是一种更为务实和稳健的策略。同时，也需要持续投入更多资源，集中力量攻克多模态数据融合、标准化以及临床验证方法学上的核心难题。

表 1：医学 AI 大模型面临的关键挑战、风险及潜在应对策略

挑战类别	具体挑战描述	对临床实践的风险	潜在缓解策略
数据问题	高质量标注数据缺乏；数据孤岛；数据隐私与安全；训练数据偏见 ²	诊断/治疗不准确；加剧健康不平等；违反法规；公众信任受损	构建多样化、代表性的训练数据集；采用联邦学习、差分隐私等技术；制定严格的数据治理和共享政策；开发偏见检测与缓解算法 ²³
模型可靠性	泛化能力不足，在真	误诊、漏诊；不当治	使用外部独立数据集

	实世界性能下降； LLM“幻觉”，生成错误信息；算法偏见 ⁴⁹	疗建议；患者伤害； 特定人群的系统性不利影响	进行严格验证；开发更鲁棒的模型架构； 采用 RAG 技术减少幻觉；持续监测和修正算法偏见；进行不确定性量化 ¹⁹
可解释性与透明度	模型“黑箱”特性，决策过程不透明 ⁴⁹	临床医生难以信任和批判性评估 AI 建议； 错误分析和调试困难；阻碍知情同意	应用 XAI 技术（如 Grad-CAM, LIME, SHAP）；透明化报告模型方法、数据和性能；在可能情况下优先选择可解释模型 ⁴⁹
临床验证与评估	缺乏临床相关的基准和大规模前瞻性临床试验；评估指标与临床实用性脱节 ¹²	难以评估 AI 的真实临床准备度和安全性； 可能部署无效或不安全的工具；阻碍比较效果研究	开发临床意义明确的验证框架和基准；强制要求严格的前瞻性试验；制定标准化报告指南（如 TRIPOD-LLM）；多中心合作 ¹²
临床整合与互操作性	难以将 AI 工具与现有 EHR 系统和临床工作流程无缝集成 ⁵⁰	采用率低；工作流程中断；效率增益有限；临床医生使用不便	开发标准化 API；加强 AI 开发者与 EHR 供应商的合作；进行以用户为中心的人机交互界面设计；提供充分的培训和技术支持 ⁵⁰
伦理、法律与社会影响 (ELSI)	责任归属不清；患者安全风险；数据治理与知情同意；对医患关系的影响；潜在的	医疗差错发生时难以追责；法律不确定性；侵犯患者自主权和隐私；医疗过程去	制定清晰的法律法规和行业指南；明确各方责任；加强伦理审查和监管；对临床医

	就业冲击 ²⁶	人化；社会接受度问题	生进行 AI 伦理培训；促进公众对话和参与 50
监管框架滞后	现有监管难以适应 AI (尤其是自适应算法和 LLM) 的快速发展和持续学习特性 ²⁶	创新受阻或监管不足导致风险；市场准入不确定性	发展适应性监管科学方法 (如预定变更控制计划 PCCP)；加强国际监管协调；建立上市后持续监测机制；鼓励沙盒监管等创新模式 ⁵²

四、医学专家大模型发展历程中的关键里程碑

医学专家大模型的发展并非一蹴而就，而是经历了一个从早期概念探索到当前技术爆发的漫长演进过程。这一历程中，关键的技术突破、理念革新以及标志性模型的出现，共同构成了其发展的里程碑。其进化路径可以被看作是从最初的“辅助感知”，逐步发展到“增强认知”，并最终朝着与人类专家“协同决策”的方向迈进。

4.1 早期探索：计算机辅助诊断（CAD）的兴起与演变

计算机在医学诊断中的应用探索可以追溯到上世纪中叶。

- **20世纪50年代**：放射学家 Lusted 首次提出了利用计算机分析放射影像中异常表现的潜力，为后续研究埋下了种子⁶。
- **20世纪60-70年代**：研究人员开始尝试将计算机技术应用于图像分析，旨在实现异常的自动检测或分类。然而，受限于当时计算机的处理能力低下和图像数字化设备的质量不高，这些早期尝试的准确性和临床可接受性均难以达到实用水平⁶。
- **20世纪80年代中期**：一个关键的理念转变发生。芝加哥大学 Kurt Rossmann 实验室的医学物理学家和放射科医生团队，包括 Giger、Doi、MacMahon 等先驱，开始大力推动计算机辅助检测（CADe）和计算机辅助诊断（CADx）的研究⁶。与早期追

求完全自动化的计算机诊断不同，他们的目标是将计算机的分析结果作为一种“辅助工具”或“第二意见”提供给放射科医生，最终的诊断决策权仍掌握在医生手中。这一转变极大地提高了 CAD 系统的临床可接受性。他们最初的研究重点是胸部 X 光片中的肺结节检测和乳腺 X 线片中的微钙化点检测⁶。Chan 等人在乳腺微钙化点自动检测方面也做出了重要贡献⁶。

- **1998 年**：首个用于乳腺 X 线筛查的商业化 CADe 系统获得了美国食品药品监督管理局（FDA）的批准，这标志着 CAD 技术开始正式进入临床应用⁶。此后，CAD 的研究和应用迅速扩展到其他多种疾病（如肺癌、结肠癌等）和不同的影像模态（如 CT、MRI、超声等）⁶。

这一时期的 CAD 系统，主要扮演的是“辅助感知”的角色。它们通过算法增强图像特征或提示可疑区域，帮助放射科医生更有效地“看片”，减少因疲劳或疏忽导致的漏诊，但其认知和决策能力有限。

4.2 技术革新：深度学习在医学影像领域的突破性应用

进入 21 世纪 10 年代，深度学习技术的崛起为医学影像分析带来了革命性的变化。

- **深度学习的赋能**：与传统的基于手工设计特征的机器学习方法不同，深度学习模型，特别是卷积神经网络（CNN），能够从大规模影像数据中自动学习和提取与诊断相关的复杂特征。这极大地提升了医学影像在分类、分割、检测等任务上的性能¹。
- **U-Net 架构的里程碑意义**：2015 年提出的 U-Net 架构，因其巧妙的编码器-解码器结构和跳跃连接设计，能够有效地融合多尺度特征信息，在医学图像（尤其是生物医学图像）分割任务中取得了巨大成功，成为该领域引用最广泛、影响最深远的模型之一，并催生了众多改进型架构²⁷。
- **Vision Transformer (ViT) 的引入**：近年来，借鉴自然语言处理领域 Transformer 模型的成功，研究者开始将 ViT 等基于 Transformer 的模型应用于医学影像分析。ViT 通过将图像分割成块并利用自注意力机制捕捉全局上下文信息，在某些任务中展现出超越传统 CNN 的潜力²⁹。
- **AI 性能比肩专家**：在一些特定的影像诊断任务中，如特定类型癌症的早期筛查（例

如乳腺癌、肺癌），基于深度学习的 AI 模型的表现开始达到甚至在某些指标上超越经验丰富的人类专家的平均水平。这极大地推动了 AI 辅助诊断工具的临床应用研究和商业化进程¹。

深度学习的应用，使得医学 AI 在“辅助感知”能力上得到空前强化，并且开始具备一定的“初级认知”能力，例如对病灶进行自动分类和风险分级。但其决策过程往往仍被视为“黑箱”，缺乏透明度和可解释性。

4.3 新范式：医学大型语言模型的诞生与迭代

如果说深度学习革新了医学影像分析，那么大型语言模型（LLMs）的出现则为医学知识的理解、处理和应用开辟了全新的范式。

- **Transformer 架构奠定基础**：2017 年谷歌提出的 Transformer 架构，凭借其高效的并行处理能力和对长距离依赖关系的捕捉，成为现代 LLM 的核心基石⁹。
- **BERT 及其医学领域变体**：2018 年发布的 BERT (Bidirectional Encoder Representations from Transformers) 模型，通过双向编码和掩码语言模型预训练，显著提升了 NLP 任务的基准水平。随后，研究者们通过在生物医学文献（如 PubMed 摘要和全文）和临床笔记上进行二次预训练或微调，开发出了一系列针对医学领域的 BERT 变体，如 BioBERT (2019 年)⁵⁵ 和 ClinicalBERT (2019 年)⁵⁷。BioBERT 在生物医学命名实体识别（NER）、关系提取等任务上表现优于通用的 BERT 模型⁵⁵，而 ClinicalBERT 则更侧重于对临床笔记的建模和下游预测任务（如预测患者再入院风险）⁵⁸。这些模型的出现，标志着预训练语言模型开始系统性地应用于医学文本处理。
- **GPT 系列与通用能力的展现**：OpenAI 开发的 GPT (Generative Pre-trained Transformer) 系列模型，特别是从 GPT-2 开始，尤其是 GPT-3 之后，展示了 LLM 在零样本/少样本学习、复杂文本生成、以及多任务处理方面的强大通用能力。研究人员也开始在医学问答、临床对话等场景测试这些通用 LLM 的性能⁹。
- **专用医学 LLM 的涌现**：为了更好地适应医学领域的专业性和严谨性，专门针对医学数据进行训练和优化的 LLM 应运而生。谷歌的 Med-PaLM (2022 年底首次预印)

及其升级版 Med-PaLM 2（2023 年发布）是其中的杰出代表⁴。这些模型通过在海量医学文本上进行指令微调和多任务学习，在回答美国执业医师资格考试（USMLE）风格的问题等医学知识问答任务上，取得了与人类专家相当甚至超越的成绩。例如，Med-PaLM 2 在 MedQA 数据集（USMLE 风格问题）上的准确率达到 86.5%⁴。其他重要的医学 LLM 还包括佛罗里达大学开发的 GatorTron 系列（2022 年发布），其最大模型参数规模达到 89 亿，专注于从电子病历（EHR）的非结构化临床笔记中解锁有价值的患者信息⁵。此外，还有基于开源 LLM（如 LLaMA）针对特定医疗场景（如医患对话）进行微调的模型，例如 ChatDoctor⁵⁹。

医学 LLM 的出现，标志着 AI 在医学领域的能力从“辅助感知”和“初级认知”向“增强认知”的重大转变。这些模型能够理解复杂的医学语言、进行一定程度的知识推理、回答专业问题，开始模拟人类医生的部分高级思维过程。

4.4 关键赋能技术：Transformer 架构与检索增强生成（RAG）等

医学专家大模型的快速发展，离不开一系列关键赋能技术的支撑。

- **Transformer 架构**：如前所述，其核心的自注意力（Self-Attention）机制使得模型能够动态地权衡输入序列中不同部分的重要性，有效捕捉长距离依赖关系。这不仅是 LLM 成功的关键，也为 Vision Transformer 等先进视觉模型提供了核心驱动力²⁹。
- **检索增强生成（Retrieval-Augmented Generation, RAG）**：为了解决 LLMs 可能产生“幻觉”（即生成不准确或虚假信息）的问题，并提高其回答的可靠性和可追溯性，RAG 技术应运而生。RAG 的核心思想是在 LLM 生成答案之前，先从一个可信的外部知识库（如医学文献数据库、临床指南、专业知识图谱等）中检索与用户提问最相关的信息片段，然后将这些检索到的信息作为上下文（context）与原始提问一起输入给 LLM，引导其生成更准确、更具事实依据的答案²¹。RAG 不仅能显著提升 LLM 在专业领域（如医疗）应用的表现，还能使其利用最新的、动态更新的知识，这对于知识快速迭代的医疗领域尤为重要³⁵。
- **其他关键技术**：此外，自监督学习（Self-supervised Learning）使得模型能够从未标注的大规模数据中学习有用的表示；多任务学习（Multi-task Learning）使模型能

够同时学习和优化多个相关任务，提升泛化能力；提示工程（Prompt Engineering）则研究如何设计有效的输入提示来引导 LLM 产生期望的输出。这些技术都对医学大模型的训练和应用起到了重要的推动作用。

当前，随着多模态融合技术的发展⁹、可解释 AI（XAI）的进步⁴⁹，以及 AI 与临床工作流程的日益深度集成²²，医学专家大模型正朝着与人类专家“协同决策”的方向发展。在这个阶段，AI 不再仅仅是提供孤立的信息或初步的判断，而是作为一种智能伙伴，与医生共同参与到复杂的临床决策过程中，提供基于证据的建议、进行风险评估、辅助制定个性化治疗方案。这种从“辅助感知”到“增强认知”再到“协同决策”的进化路径，清晰地反映了 AI 在医学领域角色的不断深化和能力的持续拓展。理解这一进化路径有助于我们把握医学 AI 的未来发展方向——未来的“医学专家大模型”将不仅仅是高效的工具，更是能够与人类医生深度协作、共同提升医疗质量和效率的智能伙伴，这对 AI 系统的设计理念、人机交互模式的创新以及未来医生的角色定位都将产生深远的影响。

值得注意的是，在医学专家大模型的发展历程中，开源与闭源模型的生态博弈也对其产生了重要影响。里程碑式的模型中，既有由大型商业公司主导开发的闭源模型（如 GPT 系列的部分版本、以及明确指出代码和权重不开源的 Med-PaLM⁴），也有由学术机构或开源社区推动发展的模型（如早期的 BERT⁵⁵、被用于 ChatDoctor 二次开发的 LLaMA 的部分版本⁵⁹，以及众多研究性质的模型）。有观点明确指出，“医学领域 LLM 的未来必须基于透明和可控的开源模型”⁶⁰，因为模型的开放性使得医疗工具开发者能够更好地控制其安全性和质量，同时也允许医疗专业人员对这些模型进行有效的监督和问责。当前，顶尖的闭源模型（如 GPT-4、Gemini 1.5 Pro⁷）通常在通用性能上表现领先，但其内部工作机制、详细的训练数据构成以及模型更新和迭代的具体过程往往不透明，这在医疗这种对安全性、可靠性和可复现性要求极高的高风险领域引发了普遍担忧⁶⁰。模型的行为甚至可能在用户不知情的情况下随时间发生显著变化，而缺乏充分的解释⁶⁰。相比之下，开源模型（如 Llama 系列的部分版本、Phi-3.5⁷）提供了更大的透明度和可定制性，允许研究者和开发者根据特定的医疗场景需求进行深度微调和优化。例如，GatorTron⁵就是一个基于学术研究但参数规模达到业界领先水平的例子，尽管其部分训练数据也来源于机构

内部的非公开临床数据。这种开源与闭源之间的生态博弈正在深刻塑造医学 AI 的发展格局。一方面，解决复杂的医学问题需要顶尖的 AI 性能；另一方面，医疗应用的特殊性（如对安全性、可靠性、可追溯性、公平性的极致要求）对模型的透明度和可控性提出了前所未有的高标准。医学领域如何在充分利用先进 AI 能力和确保其安全可信之间取得精妙平衡，将深刻影响其未来的技术选型路径和整体发展策略。未来，我们可能会看到更多针对医疗领域、经过严格验证和监管的“可信开源模型”或“半开放模型”（例如，提供开放 API 但核心模型仍受控管理的模式）的出现，以及针对闭源模型在医疗应用中更为严格和细致的审计、验证及持续监控机制的建立。

五、医学专家大模型的产业化实践与案例剖析

医学专家大模型的理论研究和技术突破，正逐步转化为实际的商业产品和解决方案，并在全球范围内推动医疗行业的智能化转型。产业化实践主要集中在医学影像 AI 和大型语言模型（LLM）在医疗领域的应用两大方向。

表 2：部分医学 AI 产业化公司及其产品/解决方案案例

公司名称 (Company)	旗舰产品/解 决方案 (Flagship Product/Solu tion)	主要应用领 域 (Main Application Area)	关键技术 (Key Technology)	主要监管批 准 (Key Regulatory Approvals)	价值证明/案 例亮点 (Value Proposition/ Case Highlights)
推想医疗 (Infervision)	InferRead® CT Lung AI	肺结节检测、 肺癌筛查、肺 密度分析	深度学习	FDA 510(k), NMPA, CE ³⁷	支持 4mm 大 小结节检测， 提供肺密度分 析及与既往影 像比对功能， 辅助早期诊断 ³⁷ 。

Paige.ai	Paige Prostate Detect, Paige PanCancer Detect	前列腺癌病理辅助诊断, 泛癌种病理检测	深度学习 (数字病理 AI)	FDA 批准 (Prostate Detect), FDA 突破性设备认定 (PanCancer Detect) ³⁸	Paige Prostate Detect 是首个获 FDA 批准的病理 AI 应用 ³⁸ 。辅助病理医生提高诊断效率和一致性。
Nanox AI (原 Zebra Medical Vision)	HealthCCSng, 其他影像 AI 算法	心血管风险评估 (冠状动脉钙化 CAC 量化), 骨质疏松、压缩性骨折 AI 检测	深度学习	FDA 510(k) (HealthCCSng, 骨质疏松等) ⁶²	HealthCCSng 通过 CT 扫描量化 CAC, 辅助心血管疾病风险评估 ⁶³ 。Nanox 整合 AI 能力, 推动数字 X 射线系统发展 ⁶² 。
联影智能 (United Imaging Intelligence)	uMI Panorama 系列, uMR Jupiter, 多款 AI 赋能影像设备	覆盖 CT, MRI, PET-CT, DR 等多种影像模态的 AI 辅助诊断	深度学习, AI 平台	FDA (超 20 款 AI 设备, 49 项 510k), NMPA, CE (uMR Jupiter 等) ⁶⁴	行业领先的 AI 赋能影像设备组合, uMR Jupiter 为全球唯一获三大主要市场批准的全身临床超高场 MRI ⁶⁴ 。
Nuance (微软旗下)	DAX™ Copilot (原	临床文档自动化 (环境临床	LLM, 语音识	HIPAA 合规 (作为解决方	自动记录医患对话并生成临

	DAX Express)	智能 ACI)	别, NLP	案) ¹⁴	床摘要, 显著减少医生文档时间 (UMH-West 每月多看 12 病人, ROI 80% ; Valley View 每接诊增加 0.1-0.3 wRVU) , 降低倦怠感 ⁴¹ 。
Abridge	Abridge Scribe	临床文档自动化 (AI 医疗书记员)	LLM, 语音识别, NLP	HIPAA 合规, 与 Epic 等 EHR 集成 ⁴⁰	实时转录医患对话并生成结构化临床笔记。MGB 试点显示笔记时间从 90 分钟减至 30 分钟以下, 39% 医生倦怠感降低 ¹⁸ 。UChicago Medicine 试点改善患者体验 ⁴⁷ 。
Epic Systems	Epic EHR AI Suite (包括 MyChart 消息回复助手, 笔记生成/摘要, 对话式 AI 等)	EHR 系统内的智能助手, 临床工作流程优化, 医患沟通	生成式 AI (LLM), 机器学习	HIPAA 合规 (作为 EHR 功能模块) ⁴⁵	约 2/3 Epic 用户已使用其 GenAI 功能, 报告显著节省行政时间 ⁴⁵ 。125 个 AI 用例在开发或

					已上线，旨在提升效率、改善体验 ¹⁷ 。 Mayo Clinic 使用其消息回复工具，每条消息节省护士约 30 秒 ¹⁷ 。
Glass Health	Glass AI Clinical Decision Support (CDS)	临床决策支持 (辅助鉴别诊断 DDx, 评估与计划 A&P)	LLM	(作为 CDS 工具, 具体监管视应用场景) 43	帮助临床医生起草鉴别诊断和诊疗计划, 结合 LLM 和医生维护的临床指南 ³ 。
其他 LLM 相关公司 (e.g., ScienceIO, emtelligent, Anthropic, GenHealth.ai)	医疗数据结构化, 临床摘要, 治疗计划生成, 医疗事件预测, 行政任务自动化等服务 ⁶⁵	覆盖医疗信息处理、临床支持、运营管理等多个方面	LLM, NLP, 机器学习	视具体产品和服务而定 (部分声称 HIPAA, SOC 2 合规) ⁶⁵	ScienceIO 将非结构化医疗文本转为结构化数据 ; emtelligent 提供医疗 AI 语言解决方案 ; Anthropic 的 Claude AI 用于自动化任务 ; GenHealth.ai 的 LMM 预测医疗事件 ⁶⁵ 。

5.1 医学影像 AI 的商业化图景

医学影像 AI 是医疗 AI 领域中商业化起步较早、发展相对成熟的分支。

领先企业及其解决方案：

如表 2 所示，多家国内外企业在医学影像 AI 领域取得了显著进展。例如，中国的推想医疗 (Infervision)，其旗舰产品 InferRead® CT Lung AI 专注于肺结节的智能检测和肺癌的辅助筛查，不仅能够自动识别小至 4 毫米的肺部结节，还具备肺密度分析、与既往影像自动比对以追踪结节变化等高级功能，已获得中国国家药品监督管理局 (NMPA)、美国食品药品监督管理局 (FDA) 的 510(k) 许可、以及欧洲 CE 认证等多国市场准入³⁷。

专注于数字病理 AI 的 Paige.ai，其 Paige Prostate Detect 是全球首个获得 FDA 批准的用于病理诊断的 AI 应用，能够辅助病理医生在数字化的前列腺活检图像中检测癌细胞，提高诊断效率和一致性³⁸。该公司还因其能够检测多种不同解剖部位癌症的 AI 应用 Paige PanCancer Detect 获得了 FDA 的“突破性设备”认定，显示了其在泛癌种 AI 诊断领域的潜力³⁸。

Zebra Medical Vision (现已被 Nanox 收购，成为 Nanox AI 的一部分) 开发了覆盖多个临床领域的医学影像 AI 算法。其 HealthCCSng 工具，可以通过分析非门控 CT 扫描图像来自动量化冠状动脉钙化 (CAC) 积分，辅助评估心血管疾病风险，并已获得 FDA 510(k) 批准⁶³。Nanox 公司自身也致力于开发创新的数字 X 射线成像系统，并计划将其与 AI 能力深度整合，以提供更智能的影像解决方案⁶²。

另一家中国企业联影智能 (United Imaging Intelligence)，在 AI 赋能的医学影像设备和解决方案方面处于行业领先地位。截至 2024 年底，该公司已有超过 20 款 AI 驱动的医疗设备获得 FDA 批准，49 款产品通过了 FDA 510(k) 认证程序，另有 46 款产品获得了欧洲 CE 认证⁶⁴。其高端产品如 uMR Jupiter，是全球唯一一款同时获得 NMPA、FDA 和 CE 三大主要市场监管机构批准的全身临床超高场 MRI 系统，体现了其在高端影像设备与 AI 技术融合方面的强大实力⁶⁴。

此外，国际巨头如 GE Aerospace (原 GE Healthcare)、Siemens Healthineers、IBM，以及一些创新型企業如 Gleamer、AZmed、Arterys、Agfa-Gevaert 等，也都在医学影像 AI 市场积极布局，推出了各自的解决方案，覆盖从影像采集优化、图像后处理增强到

计算机辅助诊断等多个环节²⁵。

市场准入与监管审批：

医学影像 AI 产品作为医疗器械或医疗器械软件（SaMD），其上市销售和临床应用必须经过严格的监管审批。不同国家和地区设立了各自的监管路径和标准。

- **美国 FDA**：AI/ML 驱动的医疗设备通常通过 510(k)（证明与已上市器械实质等同）、De Novo（针对无已上市等同器械的创新中低风险设备）或 PMA（上市前批准，针对高风险 III 类设备）途径进行审批。绝大多数已获批的 AI 影像设备是通过 510(k) 途径，且被归类为 II 类（中度风险）设备²⁶。
- **中国 NMPA**：对于新型 AI 软件，特别是没有已批准同类产品的，NMPA 可能会将其划分为风险等级较高的 III 类医疗器械进行管理，要求更为严格的临床试验和审评²⁶。
- **欧洲 CE 认证**：在欧盟地区，AI 放射学软件通常作为医疗器械进行监管，需要获得 CE 标志方可上市²⁶。

这些监管路径和标准的差异，可能导致 AI 解决方案在全球不同市场的部署进度不一，并且通常需要企业针对特定地区的要求进行额外的验证工作²⁶。例如，中国 NMPA 通常要求对每一款进口或国产的 AI 医疗器械产品进行本地化的临床试验以验证其安全性和有效性，而美国 FDA 和欧盟在一定条件下可能认可国外的临床研究数据²⁶。

尽管面临复杂的监管环境，全球范围内获得批准的 AI/ML 医疗设备数量仍在快速增长。据统计，截至 2023 年 10 月，FDA 已批准超过 690 种 AI 赋能的医疗设备，其中高达 77% 应用于放射学领域⁵³。另一份报告指出，到 2025 年初，FDA 批准的临床 AI 算法已达 1000 种，其中 76% 同样服务于放射科¹⁰。这充分说明放射学是当前医学 AI 应用最为集中和成熟的领域。

5.2 大型语言模型在医疗产业的应用

相较于医学影像 AI，大型语言模型（LLMs）在医疗产业的应用虽然起步稍晚，但凭借其在处理非结构化文本、理解复杂语境和生成自然语言方面的强大能力，正以惊人的速度渗

透到医疗的各个环节，展现出巨大的商业化潜力。

临床文档与报告自动化：

这是目前 LLM 在医疗领域最成功、最具影响力的应用之一，旨在将医生从繁重的临床文书工作中解放出来。

- **Nuance DAX Copilot**：作为微软旗下的 Nuance 公司推出的环境临床智能 (Ambient Clinical Intelligence, ACI) 解决方案，DAX Copilot (前身为 DAX Express) 能够在医患交流过程中，通过麦克风实时捕捉对话内容，并利用 LLM 和先进的语音识别、自然语言处理技术，自动将其转换为结构化的临床摘要或病历草稿，然后无缝集成到主流的电子病历 (EHR) 系统中¹⁴。多项案例研究表明，DAX Copilot 能够显著减少医生在文档记录上花费的时间，有效降低职业倦怠感。例如，密歇根大学健康西部医院 (University of Michigan Health-West) 的报告显示，使用 DAX Copilot 的医生每月能够多看 12 名患者，并且实现了高达 80% 的投资回报率 (ROI)⁴¹。Valley View Hospital 的案例也指出，DAX Copilot 帮助医生平均每次接诊增加了 0.1 至 0.3 个工作相对价值单位 (wRVU)，这直接反映了其对工作效率和经济效益的积极影响⁴²。
- **Abridge**：是另一家在 AI 医疗“书记员”领域领先的公司。Abridge 平台同样能够实时转录医患对话，并利用生成式 AI 技术将其转化为结构化的临床笔记，支持与 Epic 等主流 EHR 系统的集成⁴⁰。在麻省总医院布莱根 (Mass General Brigham) 进行的试点项目中，Abridge 帮助医生将撰写临床笔记的时间从平均 90 分钟大幅缩短至 30 分钟以下，并且有 39% 的参与医生表示其职业倦怠感有所降低¹⁸。芝加哥大学医学中心的试点也显示，Abridge 的应用改善了患者的就医体验，因为医生能够更专注于与患者的交流⁴⁷。
- 其他类似解决方案还包括 **DeepScribe** 等，它们共同构成了快速增长的 AI 临床文档自动化市场²²。

集成于电子病历 (EHR) 系统的智能助手：

EHR 系统是现代医疗信息化的核心。将 LLM 的智能能力嵌入 EHR，能够极大地提升其易用性和功能性。

- **Epic Systems**：作为美国市场份额最大的 EHR 供应商，Epic 正积极地将其 EHR 系统与 AI 技术，特别是生成式 AI，进行深度融合⁴⁵。Epic 开发的 AI 功能模块包括：帮助患者准备就诊的对话式 AI（例如，在就诊前询问患者就诊目的、是否需要安排检查等）；集成在 MyChart 患者门户中的医患消息自动回复草稿功能，能够根据患者提问和病历信息预生成回复建议；临床笔记的自动生成与智能摘要；辅助医生进行医嘱录入和临床路径推荐等¹⁶。据报道，已有约三分之二的 Epic 用户开始使用其提供的生成式 AI 功能，并普遍反映这些功能显著节省了他们的行政工作时间⁴⁵。Epic 目前已有 125 个 AI 相关的应用案例正在开发或已经上线，其目标是全面提升临床工作流程效率、改善医患交互体验¹⁷。例如，梅奥诊所（Mayo Clinic）使用 Epic 集成的消息回复助手后，护士处理每条患者消息的时间平均节省了约 30 秒，并且 AI 生成的回复草稿在同理心表达方面也表现良好¹⁷。
- 其他主要的 EHR 厂商，如 **Oracle Health (原 Cerner)**、**Meditech**、**eClinicalWorks** 等，也都在积极探索和整合 AI 技术，以提升其产品的智能化水平和市场竞争力⁴⁶。

临床决策支持（CDS）产品：

LLMs 在辅助医生进行复杂临床决策方面也展现出潜力。

- **Glass Health**：该公司提供了一个 AI 驱动的临床决策支持平台，旨在帮助临床医生根据患者的症状和病史，快速起草鉴别诊断（Differential Diagnosis, DDx）列表，并生成初步的评估与治疗计划（Assessment & Plan, A&P）草案，供医生参考和完善³。Glass Health 的特点是将其大型语言模型与由执业医师团队维护和更新的临床指南相结合，以确保其建议的循证医学基础⁴⁴。
- 其他 LLM 在 CDS 中的应用案例还包括，纽约大学朗格尼健康中心（NYU Langone Health）与 NVIDIA 合作开发的 NYUTron 模型，该模型能够分析患者的电子病历数据，预测其在出院后 30 天内再次入院的风险，从而帮助医疗团队识别高风险患者并提前进行干预³。

其他商业化探索与产品：

除了上述主要应用方向外，LLMs 在医疗产业的商业化探索还体现在：

- **医疗数据处理与分析服务**：一些公司如 **ScienctelO**、**emtelligent**、**Anthropic** (其 **Claude AI** 模型)、**GenHealth.ai**、**Gradient AI** 等，利用 LLM 技术提供多样化的医疗信息服务，包括将非结构化的医疗文本（如病历、报告）自动转换为结构化数据，生成临床摘要，辅助制定个性化治疗计划，预测未来医疗事件（如疾病风险、医疗费用），以及自动化处理各类医疗行政任务（如医疗编码、账单审核、患者日程安排）等⁶⁵。
- **药物研发与临床试验**：如前文所述，**Exscientia** 等公司已成功将 AI（包括 LLM 技术）应用于新药发现的早期阶段，加速候选化合物的筛选和设计³。
- **AI 聊天机器人与虚拟助手**：用于患者健康咨询、慢性病管理支持、心理健康初步干预等场景的 AI 聊天机器人也逐渐出现，它们能够提供个性化的信息和支持，改善患者的自我管理能力和就医体验³。

5.3 全球市场规模、增长预测与投资趋势

医学 AI（包括影像 AI 和基于 LLM 的应用）市场正处于高速增长阶段，吸引了大量投资。

- **医学影像 AI 市场规模**：根据 Grand View Research 的报告，全球医学影像 AI 市场规模在 2022 年约为 7.54 亿美元，预计到 2030 年将增长至 81.778 亿美元，期间的复合年增长率（CAGR）高达 34.7%²⁵。另一份来自 Precedence Research 的报告则预测，该市场规模在 2024 年为 12.8 亿美元，到 2034 年有望达到 144.6 亿美元，预测期（2025-2034）的 CAGR 为 27.10%⁴⁵。尽管具体的数字和预测期略有差异，但两份报告均清晰地指出了医学影像 AI 市场未来强劲的增长势头。
- **技术细分市场趋势**：在医学影像 AI 的技术构成中，深度学习是当前占据最大市场份额的技术细分。而从增长潜力来看，自然语言处理（NLP）技术（LLM 是其核心）预计将成为增长最快的细分市场²⁵。这反映了 LLM 在医疗领域应用的广阔前景和快速发展态势。
- **区域市场格局**：目前，北美地区（尤其是美国）是全球医学影像 AI 最大的收入市场，并且预计在未来几年内将继续引领全球市场的收入增长²⁵。

- **投资趋势**：医疗机构和生物制药公司对 AI 技术的投资意愿强烈。一份报告指出，60%的生物制药企业高管已经为其 AI 相关的研发项目设定了近期的投资回报 (ROI) 目标¹³。医疗机构也期望通过引入 AI 来改善患者的临床预后，同时提高运营效率和成本效益¹³。

当前医学 AI 的产业化格局呈现出一种清晰的态势：医学影像 AI 凭借其在特定诊断任务上的明确效用、相对结构化的数据处理以及深度学习技术的早期突破，率先实现了较为成熟的商业化。市场上已有多款影像 AI 产品获得了主流监管机构（如 FDA、NMPA、CE）的批准，并已在临床实践中得到应用，其核心价值主要体现在提升诊断的效率和准确性方面²⁶。

与此同时，大型语言模型（LLM）在医疗领域的应用虽然起步稍晚，但正以惊人的速度追赶，并致力于拓展 AI 在医疗健康领域的价值边界。尤其是在临床文档自动化²²、EHR 系统智能化集成⁴⁵ 以及临床决策支持⁴³ 等方面，LLM 的应用发展势头尤为迅猛。这主要得益于近年来 LLM 技术的爆发式进步及其在处理和理解复杂非结构化文本方面的独特优势。LLM 的应用不仅能够优化现有的医疗流程（例如，通过自动化文档记录来大幅减少医生的文书工作时间），更重要的是，它还在开辟全新的价值创造领域，例如改善医患之间的沟通与互动³、赋能药物的早期发现与研发²⁰、提供个性化的患者教育与健康管理支持³ 等，这些往往是传统影像 AI 技术难以直接覆盖的场景。市场预测数据也间接印证了这一趋势，自然语言处理（NLP）被认为是未来几年医疗 AI 中增长最快的技术细分领域²⁵。因此，可以预见，虽然目前影像 AI 在特定诊断领域已形成了较为成熟的商业模式，但 LLM 凭借其更广泛的适用性和更深层次的认知赋能潜力，其应用的广度和深度有望超越传统影像 AI，为医疗健康行业带来更为广泛和深刻的价值创造。未来，影像 AI 与 LLM 的深度融合，例如基于多模态信息（结合影像特征和文本报告）的智能诊断、影像报告的自动生成与解读等，将是重要的技术创新方向和产业增长点。

在医学 AI 产品走向商业化的过程中，获得监管机构的批准（如 FDA、NMPA、CE 认证等）无疑是进入市场的关键门槛和先决条件²⁶。没有相应的合规资质，AI 产品无法在临床环境中合法销售和使用。然而，仅仅获得监管批准并不等同于商业上的成功。产品能

否在真实世界的临床环境中持续创造并证明其独特的临床价值，以及能否无缝地集成到医疗机构现有的复杂工作流程中，才是其能否获得市场认可并实现持续成功的核心要素。临床价值的证明需要依赖于严谨的临床试验结果³⁶ 和上市后积累的真实世界研究数据，这些数据需要清晰地展示 AI 工具如何能够实际改善诊疗效果（如提高诊断准确率、降低并发症）、提升工作效率（如减少医生工作负荷、缩短患者等待时间¹⁶）、改善患者体验或最终降低医疗成本。同时，AI 工具必须能够便捷地嵌入到医生日常使用的电子病历（EHR）系统、影像归档和通信系统（PACS）等现有信息基础设施中¹⁰，而不是给临床医生增加额外的操作负担或割裂原有的工作习惯。EHR 的无缝集成⁴⁰ 是一个典型的例子，它直接影响到 AI 工具的易用性和采纳率。如果一个 AI 工具虽然技术先进，但在实际使用中操作不便，或者与医疗机构现有的信息系统不兼容，那么即使它通过了监管审批，也很难被临床医生广泛和持续地采用。回顾前述成功的产业化案例，如推想医疗的肺结节 AI³⁷、Paige 的病理 AI³⁸、Nuance 的 DAX Copilot⁴¹ 以及 Abridge 的 AI 书记员¹⁸ 等，它们不仅成功跨越了监管的门槛，更重要的是，它们在各自的临床应用场景中都展现了可量化的效益（例如，显著的时间节省、工作效率的提升、医生倦怠感的降低等），并且高度注重与医疗机构现有工作流程（特别是 EHR 系统）的集成。因此，对于医学 AI 企业而言，其发展战略不仅要投入大量资源攻克核心技术难题和应对复杂的监管审批流程，更要从产品立项之初就深入洞察真实的临床需求，明确产品的价值主张，并精心设计其与现有医疗生态系统的融合方式。始终以用户（临床医生和患者）为中心，用真实的数据证明产品的临床价值和经济效益，是医学 AI 产品穿越商业化“死亡谷”、实现可持续发展的关键所在。

六、总结与前瞻

6.1 核心研究发现与洞察凝练

本报告对医学专家大模型进行了多维度的调研与分析，核心研究发现与洞察可凝练如下：

1. 定义与演进：“医学专家大模型”是一个动态发展的概念，已从早期辅助特定任务（如 CAD 辅助感知）的 AI，演进为能够在医学影像分析（增强感知与初级认知）和医学

语言处理（增强认知与推理）两大核心方向展现出高级别专业能力的综合性 AI 系统，并呈现出多模态融合的趋势。其发展是医疗领域迫切需求与 AI 技术突破双轮驱动的结果。

2. **核心价值彰显**：医学专家大模型在提升诊断准确性与效率（如癌症早筛、病理辅助）、赋能医学研究与药物发现（如加速文献回顾、优化临床试验）、优化临床工作流程与决策支持（如自动化临床文档、EHR 智能助手、CDS）以及推动个性化医疗与患者管理等方面，已展现出巨大的应用价值和系统性赋能潜力。其价值实现正从单点技术突破转向对医疗全流程的系统性赋能。
3. **学术前沿与瓶颈**：学术研究在前沿模型架构（如 Transformer、U-Net 变体、专用医学 LLM）、多模态融合、可信 AI（XAI、RAG）、数据高效利用与隐私保护等方面持续探索。然而，模型发展仍面临数据获取与治理困境、泛化能力与临床适用性差距、“幻觉”与算法偏见、临床验证复杂性与标准化评估缺失，以及滞后的伦理法规监管等核心挑战。
4. **信任是核心**：在所有技术、流程和伦理瓶颈中，“信任”是贯穿始终的核心障碍。建立临床医生、患者和监管机构对 AI 系统可靠性、安全性和公平性的信任，是医学专家大模型成功推广应用的前提。
5. **发展里程碑**：其发展历程可概括为从早期 CAD 的“辅助感知”，到深度学习影像 AI 的“增强感知与初级认知”，再到医学 LLM 的“增强认知”，并正朝着与人类专家“协同决策”的未来演进。开源与闭源模型的生态博弈也深刻影响着其发展路径。
6. **产业化格局**：产业化呈现出医学影像 AI 率先成熟，而 LLM 应用加速追赶并拓展价值边界的格局。监管审批是市场准入的关键门槛，但持续的临床价值证明和无缝的工作流整合是商业成功的核心。

6.2 未来发展趋势、潜在突破口与战略建议

展望未来，医学专家大模型的发展将呈现以下趋势，并可能在若干关键领域实现突破。为促进其健康发展，需要多方协同努力。

技术趋势展望：

- **多模态融合的深化与智能化**：未来的医学 AI 将不仅仅是简单地叠加不同来源的数据，而是能够更智能地理解和整合来自影像、文本、基因组学、蛋白质组学、可穿戴设备等多源异构医疗数据之间的复杂关联，从而实现更全面、更精准的疾病诊断、风险预测、预后评估和治疗决策支持²。
- **因果 AI 与可信 AI 的构建**：当前多数 AI 模型主要依赖于数据中的相关性进行预测。未来的研究将更加注重探索数据背后的因果关系，发展因果推断能力。结合可解释人工智能（XAI）、检索增强生成（RAG）以及更严格的验证方法，构建真正可信赖、决策过程透明且结果可溯源的医学 AI 系统，是提升临床接受度的关键⁴⁹。
- **持续学习与自适应模型的演进**：开发能够在真实临床环境中安全、有效地进行持续学习和适应新知识、新指南的模型，同时解决由此带来的监管难题（如模型的动态更新与验证），是 AI 系统保持长期临床价值的重要方向⁵³。
- **边缘计算与端侧 AI 的普及**：为了更好地保障患者数据隐私、降低网络延迟、实现实时响应，将 AI 模型的计算更多地部署在医疗设备端（如影像设备、监护仪）或医院本地服务器（边缘计算节点）将成为趋势。
- **个性化与精准化应用的深化**：AI 在个体化疾病风险预测（如基于多组学数据的癌症风险模型）、精准治疗方案选择（如根据患者特异性分子标志物推荐靶向药物）、药物反应与副作用预测等方面的应用将更加深入和精细化。

潜在突破口：

- **高质量、标准化、可共享的医学数据集的构建**：通过建立国家级或区域性的医学数据共享平台，采用联邦学习、安全多方计算等隐私保护技术，促进大规模、高质量、多样化且标准化的医学数据集的形成，将极大地推动模型研发和验证²。
- **针对 LLM“幻觉”和偏见问题的更有效缓解机制**：在 RAG 的基础上，进一步研究和开发能够更有效地检测、量化和纠正 LLM 输出中潜在错误、不实信息（“幻觉”）以及算法偏见的机制，对于确保其在医疗领域的安全应用至关重要⁵⁰。
- **建立被广泛认可的医学 AI 临床验证标准和评估平台**：形成一套科学、严谨、且被临床界和监管机构共同认可的医学 AI 产品临床验证标准、评估方法学和基准测试平

台，将有助于规范市场、加速优质产品的审评审批和推广应用¹²。

- **AI 伦理原则在模型设计、开发、部署全流程的有效嵌入**：将公平、透明、问责、隐私保护等伦理原则深度融入医学 AI 系统的整个生命周期，确保技术发展符合人类共同的价值观。
- **创新的监管科学方法与工具**：发展能够适应 AI 技术快速迭代和模型复杂性的新型监管科学方法和工具，例如“沙盒监管”、基于真实世界证据的持续监控和评估机制等，以平衡创新激励与风险管控。

战略建议（针对研究机构、产业界、政策制定者）：

- **加强跨学科合作**：大力促进人工智能技术专家、临床医学专家、数据科学家、伦理学家、法律专家以及患者代表之间的深度对话与紧密合作，确保 AI 技术的研发和应用能够真正解决临床问题并满足实际需求。
- **鼓励开发和采用开源、透明的医学 AI 工具和平台**：支持构建和共享经过严格验证的开源医学 AI 模型、数据集和开发工具，以降低研发门槛，促进技术创新和公平竞争，并增强系统的透明度和可信度⁶⁰。
- **加大对医学 AI 人才培养的投入**：培养一批既懂 AI 技术又了解医学知识的复合型人才，是推动医学 AI 持续发展的关键。
- **建立健全医学 AI 的伦理规范、法律法规和监管体系**：制定清晰、适应性强且具有前瞻性的伦理指南、法律框架和监管政策，为医学 AI 的健康发展提供明确的边界和保障，有效平衡技术创新与潜在风险。
- **推动真实世界数据（RWD）和真实世界证据（RWE）的应用**：鼓励利用在常规临床实践中收集的真实世界数据来验证和迭代 AI 模型，并探索将真实世界证据作为支持 AI 产品监管决策和临床应用的重要依据。

最终，医学专家大模型的未来发展，将经历一个从追求“工具理性”（即强调技术性能、效率提升和经济回报¹⁹）到实现“价值共创”的深刻转变。医学的本质是关乎人的生命、健康与福祉，它天然地交织着复杂的伦理、社会和情感因素。如果 AI 技术的发展仅仅停留在追求冰冷的技术指标提升，而不能真正转化为患者的实际获益、临床医生职业幸福感的增

强、以及整个医疗系统运行效率和公平性的整体优化，那么其所能创造的价值将是有限的。未来医学AI的发展趋势，无论是对多模态信息的深度融合、对可信AI的不懈追求，还是对个性化医疗的精细化探索，都清晰地指向AI需要更深刻地理解医疗场景的内在复杂性和人的多层次需求。相应的，战略层面的建议，如强调跨学科合作、将伦理原则嵌入技术实践、以及重视真实世界证据的积累与应用，也无不体现出需要AI开发者、临床医生、患者、政策制定者等所有利益相关方共同参与，共同定义和实现AI在医疗健康领域应有的价值，这即是“价值共创”的核心理念。这意味着，衡量未来医学专家大模型成功与否的标准，将远不止于其技术性能的先进性，更在于其能否在整个医疗生态系统中，为促进健康公平、提升人文关怀、优化资源配置、增强系统韧性等多个维度做出实质性的贡献。未来的医学专家大模型，不仅要“能干”，更要“好用”、“可信”、“向善”，真正成为人类医学专家的得力伙伴，共同为增进人类健康福祉而努力。

七、参考文献

- ¹ ebiotrade.com - AI/DL in cancer early detection, diagnosis, prognosis.
(<https://www.ebiotrade.com/newsf/2025-5/20250521133401324.htm>)
- ⁶⁷ ebiotrade.com - Mentions cancer research topics.
(<https://www.ebiotrade.com/newsf/2025-5/20250528074116309.htm>)
- ²⁶ ijsrtjournal.com - Regulatory pathways for AI in radiology (FDA, CE, NMPA).
(<https://www.ijsrtjournal.com/article/Artificial+Intelligence+in+Radiology+Transforming+Diagnostics+and+Raising+Ethical+Dilemmas>)
- ¹⁹ arxiv.org - Generalizable and Explainable DL for Medical Image Computing.
(<https://arxiv.org/abs/2503.08420>)
- ⁶ pmc.ncbi.nlm.nih.gov - Historical milestones of CAD.
(<https://pmc.ncbi.nlm.nih.gov/articles/PMC2673617/>)
- ⁴⁹ arxiv.org - XAI techniques in medical imaging (GradCAM, LIME, etc.).
(<https://arxiv.org/html/2503.08420v1>)
- ⁶ pmc.ncbi.nlm.nih.gov - Key historical milestones of CAD development.
(<https://pmc.ncbi.nlm.nih.gov/articles/PMC2673617/>)
- ² pmc.ncbi.nlm.nih.gov - Deep Learning for Medical Image-Based Cancer Diagnosis review. (<https://pmc.ncbi.nlm.nih.gov/articles/PMC10377683/>)
- ³⁷ infervision.com.tr - Infervision InferRead CT Lung FDA 510(k) clearance.
(<https://www.infervision.com.tr/en/news/infervision%2C-inferread-ct-akci%C4%9Ferindeki-geli%C5%9Fmi%C5%9F-%C3%B6zellikler->)

- i%CC%87%C3%A7in-fda-510(k)-onay%C4%B1-ald%C4%B1
- ² pmc.ncbi.nlm.nih.gov - Deep learning in medical image-based cancer diagnosis. (<https://pmc.ncbi.nlm.nih.gov/articles/PMC10377683/>)
 - ⁶³ mobihealthnews.com - Zebra Medical Vision (Nanox) FDA clearance for cardiac tool. (<https://www.mobihealthnews.com/news/zebra-medical-vision-lands-fda-clearance-tool-detect-cardiovascular-disease>)
 - ⁶² fiercebiotech.com - Nanox FDA clearance for X-ray bed and AI programs. (<https://www.fiercebiotech.com/medtech/nanox-scores-clearance-cloud-connected-x-ray-bed-after-years-fda-review>)
 - ²⁹ arxiv.org - Advances in Medical Image Analysis with Vision Transformers. (<https://arxiv.org/abs/2301.03505>)
 - ⁴⁹ arxiv.org - Challenges of DL in medical imaging: explainability, generalizability. (<https://arxiv.org/html/2503.08420v1>)
 - ¹⁹ arxiv.org - Overview of generalizable and explainable AI in DL for medical imaging. (<https://arxiv.org/abs/2503.08420>)
 - ¹² um.edu.mt - Clinical impact of AI in radiology, challenges, expert opinion. (https://www.um.edu.mt/library/oar/bitstream/123456789/132891/1/2519EMAEMA590705028413_1.PDF)
 - ⁶¹ global.infervision.com - Infervision InferRead CT Lung product details. (<https://global.infervision.com/products/inferread-ct-lung>)
 - ¹⁰ marshall.edu - AI in radiology: efficiency, accuracy, workload reduction. (<https://mds.marshall.edu/cgi/viewcontent.cgi?article=2948&context=etd>)
 - ²³ preprints.org - AI in cancer imaging: review covering radiomics, DL, emerging trends. (https://www.preprints.org/frontend/manuscript/11e4cc1244d3f2748f00570b992642a0/download_pub)
 - ²⁴ rsna.org - AI as a second reader in mammography, clinical trial results. (<https://www.rsna.org/news/2025/march/ai-as-a-second-reader-in-mammography>)
 - ²⁶ ijsrtjournal.com - AI in radiology: applications, benefits, shortcomings, ethics, regulation. (<https://www.ijsrtjournal.com/article/Artificial+Intelligence+in+Radiology+Transforming+Diagnostics+and+Raising+Ethical+Dilemmas>)
 - ³⁰ arxiv.org - Vision Transformers in Medical Imaging: A Review. (<https://arxiv.org/abs/2211.10043>)
 - ²⁵ grandviewresearch.com - AI in medical imaging market size and forecast. (<https://www.grandviewresearch.com/horizon/outlook/ai-in-medical-imaging->)

market-size/global)

- ²⁷ [digitalocean.com](https://www.digitalocean.com/community/tutorials/unet-architecture-image-segmentation) - U-Net architecture for medical image segmentation.
(<https://www.digitalocean.com/community/tutorials/unet-architecture-image-segmentation>)
- ¹⁵ [ramsoft.com](https://www.ramsoft.com/blog/future-of-ai-in-radiology) - Future of AI in radiology, adoption challenges, ethics.
(<https://www.ramsoft.com/blog/future-of-ai-in-radiology>)
- ⁵⁰ [pmc.ncbi.nlm.nih.gov](https://PMC12086438/) - Ethical challenges of LLMs in healthcare: bias, safety, privacy. (<https://PMC12086438/>)
- ³⁸ [paige.ai](https://www.paige.ai/press-releases/us-fda-grants-paige-breakthrough-device-designation-for-ai-application-that-detects-cancer-across-different-anatomic-sites) - Paige PanCancer Detect FDA Breakthrough Device designation.
(<https://www.paige.ai/press-releases/us-fda-grants-paige-breakthrough-device-designation-for-ai-application-that-detects-cancer-across-different-anatomic-sites>)
- ²⁸ [pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC7329239/) - UNet++ architecture for medical image segmentation.
(<https://pmc.ncbi.nlm.nih.gov/articles/PMC7329239/>)
- ⁶⁸ [globenewswire.com](https://www.globenewswire.com/news-release/2025/02/13/3026027/0/en/AI-in-Medical-Imaging-Market-Size-Projected-to-Reach-USD-14-46-Bn-By-2034.html) - AI in medical imaging market size projection.
(<https://www.globenewswire.com/news-release/2025/02/13/3026027/0/en/AI-in-Medical-Imaging-Market-Size-Projected-to-Reach-USD-14-46-Bn-By-2034.html>)
- ⁶ [pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC2673617/) - History of CAD in radiology.
(<https://pmc.ncbi.nlm.nih.gov/articles/PMC2673617/>)
- ³² [customerthink.com](https://customerthink.com/ais-symbiotic-impact-on-drug-development-and-patient-experience-in-global-pharma-and-biotech/) - AI in drug development and clinical trials.
(<https://customerthink.com/ais-symbiotic-impact-on-drug-development-and-patient-experience-in-global-pharma-and-biotech/>)
- ³⁹ [ncbi.nlm.nih.gov](https://www.ncbi.nlm.nih.gov/books/NBK608438/) - Paige Prostate Suite details.
(<https://www.ncbi.nlm.nih.gov/books/NBK608438/>)
- ³⁶ [pubs.rsna.org](https://pubs.rsna.org/doi/10.1148/ryai.240287) - AI algorithm performance in breast cancer detection on mammograms. (<https://pubs.rsna.org/doi/10.1148/ryai.240287>)
- ⁶⁹ [pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC1955762/) - CAD for lung nodules on chest radiographs.
(<https://pmc.ncbi.nlm.nih.gov/articles/PMC1955762/>)
- ³ [paubox.com](https://www.paubox.com/blog/large-language-models-in-healthcare) - LLMs in healthcare: definition, capabilities, applications.
(<https://www.paubox.com/blog/large-language-models-in-healthcare>)
- ²⁰ [jmir.org](https://www.jmir.org/2025/1/e59069/) - LLMs in medical AI: NLP, clinical support, research, challenges.
(<https://www.jmir.org/2025/1/e59069/>)
- ⁷ [hai.stanford.edu](https://hai.stanford.edu/news/holistic-evaluation-of-large-language-models-for-medical-applications) - Holistic evaluation of LLMs for medical applications.
(<https://hai.stanford.edu/news/holistic-evaluation-of-large-language-models-for-medical-applications>)
- ²¹ mdpi.com - LLMs in medical education, clinical decision support,

- administration; RAG. (<https://www.mdpi.com/2227-9032/13/6/603>)
- ⁷⁰ mdpi.com - LLM literature in medicine: temporal, geographical, thematic analysis. (<https://www.mdpi.com/1999-5903/17/2/76>)
 - ³³ multimodal.dev - Applications of LLMs in healthcare: admin, patient engagement, research. (<https://www.multimodal.dev/post/large-language-models-in-healthcare>)
 - ¹¹ amzur.com - AI in healthcare: challenges (bias, data, security, cost, integration, ethics). (<https://amzur.com/blog/ai-in-healthcare-challenges-solutions>)
 - ¹³ healthcareitnews.com - Healthcare AI adoption, data and integration challenges. (<https://www.healthcareitnews.com/news/healthcare-ai-adoption-data-and-integration-challenges-persist>)
 - ⁹ pmc.ncbi.nlm.nih.gov - LLMs in medical AI, multimodal data, challenges.²⁰ (<https://pmc.ncbi.nlm.nih.gov/articles/PMC11751657/>)
 - ³¹ ai21.com - Benefits of LLMs for clinicians, patients, researchers. (<https://www.ai21.com/knowledge/l1ms-in-healthcare/>)
 - ⁴ pmc.ncbi.nlm.nih.gov - Med-PaLM 2 Google research paper. (<https://pmc.ncbi.nlm.nih.gov/articles/PMC11922739/>)
 - ⁸ research.google - Med-PaLM overview by Google Research. (<https://sites.research.google/med-palm/>)
 - ⁵ researchgate.net - GatorTron: Large Clinical Language Model. (https://www.researchgate.net/publication/358938843_GatorTron_A_Large_Clinical_Language_Model_to_Unlock_Patient_Information_from_Unstructured_Electronic_Health_Records)
 - ⁴⁸ medrxiv.org - GatorTron preprint. (<https://www.medrxiv.org/content/10.1101/2022.02.27.22271257v2>)
 - ⁷¹ jmir.org - Accuracy of LLMs (ChatDoctor mentioned in related links) for clinical questions. (<https://www.jmir.org/2025/1/e64486>)
 - ⁵⁹ github.com - ChatDoctor: Medical Chat Model based on LLaMA. (<https://github.com/KentOn-Li/ChatDoctor>)
 - ²¹ mdpi.com - Review of LLMs in medicine.²¹ (<https://www.mdpi.com/2227-9032/13/6/603>)
 - ⁶⁰ researchgate.net - The path forward for LLMs in medicine is open (open-source models). (https://www.researchgate.net/publication/386193568_The_path_forward_for_large_language_models_in_medicine_is_open)
 - ⁵⁰ pmc.ncbi.nlm.nih.gov - Challenges of LLMs: hallucination, validation, privacy, bias, ethics (Tables 2 & 3). (<https://pmc.ncbi.nlm.nih.gov/articles/PMC12086438/>)

- ⁵¹ pmc.ncbi.nlm.nih.gov - ChatGPT in healthcare: risks, challenges, limitations. (<https://pmc.ncbi.nlm.nih.gov/articles/PMC12057020/>)
- ³⁴ pmc.ncbi.nlm.nih.gov - Challenges of RAG in medical LLMs. (<https://pmc.ncbi.nlm.nih.gov/articles/PMC12059965/>)
- ³⁵ mdpi.com - RAG for medical LLMs, enhancing transparency and reliability. (<https://www.mdpi.com/2504-4990/6/4/116>)
- ⁹ pmc.ncbi.nlm.nih.gov - History of LLMs from BERT to GPT (timeline figure mentioned). (<https://pmc.ncbi.nlm.nih.gov/articles/PMC11751657/>)
- ²⁰ jmir.org - History of LLMs, BERT, GPT.²⁰ (<https://www.jmir.org/2025/1/e59069/>)
- ³ paubox.com - Commercial medical LLMs (Epic-Microsoft, Google AIME, BioBERT, ClinicalBERT, Med-PaLM2). (<https://www.paubox.com/blog/large-language-models-in-healthcare>)
- ⁶⁵ elion.health - Commercial medical LLM products (emtelligent, ScienceO, Anthropic Claude, GenHealth.ai, Gradient AI). (<https://elion.health/categories/large-language-models/products>)
- ²² elion.health - Nuance DAX, Abridge, DeepScribe for clinical documentation. (<https://elion.health/products/nuance-dax>)
- ¹⁴ nuance.com - Nuance DAX for automated clinical documentation. (<https://www.nuance.com/healthcare/dragon-ai-clinical-solutions/dax-copilot/infographic/move-beyond-scribes-to-automatically-document-care.html>)
- ⁴⁵ advisory.com - Epic Systems AI in EHR, new capabilities. (<https://www.advisory.com/daily-briefing/2025/03/14/ep-ehr-ec>)
- ⁴⁶ healthcaredive.com - EHR vendors (Epic, Oracle) turn to AI. (<https://www.healthcaredive.com/news/ehr-artificial-intelligence-efforts-epic-oracle/741541/>)
- ⁴⁰ revmaxx.co - Abridge medical scribe review, features, EHR integration. (<https://www.revmaxx.co/abridge/>)
- ⁶⁶ scribehealth.ai - Abridge alternatives, pricing. (<https://www.scribehealth.ai/compare/abridge-alternatives>)
- ⁴³ glass.health - Glass Health AI Clinical Decision Support platform. (<https://glass.health/>)
- ⁴⁴ callin.io - Glass Health AI for diagnosis and clinical decision support. (<https://callin.io/ai-tools/glass-health/>)
- ⁵⁴ timelines.issarice.com - Timeline of large language models (general, includes BERT, Transformer). (https://timelines.issarice.com/wiki/Timeline_of_large_language_models)

- ⁹ pmc.ncbi.nlm.nih.gov - Timeline of mainstream LLMs (mentions RNN, LSTM, Transformer, BERT, GPT). (<https://pmc.ncbi.nlm.nih.gov/articles/PMC11751657/>)
- ⁴¹ dictationdirect.com - Nuance DAX Copilot case study: University of Michigan Health-West (ROI). (<https://www.dictationdirect.com/case-studies/dax-copilot-university-of-michigan-health-west/>)
- ⁴² athenahealth.com - Nuance DAX Copilot case study: Valley View Hospital (ROI). (<https://www.athenahealth.com/resources/case-studies/valley-view-hospital>)
- ¹⁶ spsoft.com - Epic EHR AI trends, generative AI implementation results (time/burnout reduction). (<https://spsoft.com/tech-insights/epic-ehr-ai-trends-in-2025-reshaping-care/>)
- ¹⁷ fiercehealthcare.com - Epic Launchpad for generative AI adoption. (<https://www.fiercehealthcare.com/ai-and-machine-learning/epic-introduces-launchpad-fuel-faster-generative-ai-adoption-among>)
- ¹⁸ massgeneralbrigham.org - Abridge AI scribe at Mass General Brigham, physician feedback. (<https://www.massgeneralbrigham.org/en/about/newsroom/articles/how-shiv-raos-abridge-is-freeing-clinicians-from-administrative-burdens>)
- ⁴⁷ mobius.md - AI-powered medical scribes hospital deployments (Stanford/DAX, UChicago/Abridge). (<https://www.mobius.md/blog/hospitals-are-embracing-ai-powered-medical-scribes>)
- ⁵² ldi.upenn.edu - FDA regulation of LLMs in medicine, challenges with generative AI. (<https://ldi.upenn.edu/our-work/research-updates/ai-in-health-care-and-the-fdas-blind-spot/>)
- ⁵³ himss.org - Navigating regulatory landscape for LLMs, FDA oversight. (<https://www.himss.org/resources/section-5-navigating-regulatory-landscape-analysis-legal-and-ethical-oversight-large>)
- ⁵⁵ scispace.com - BioBERT research paper. (<https://scispace.com/papers/biobert-a-pre-trained-biomedical-language-representation-4lpqv1ete>)
- ⁵⁶ neuralmagic.com - Sparse BioBERT for biology research. (<https://neuralmagic.com/blog/revolutionizing-biology-research-with-lightning-fast-nlp-introducing-sparse-biobert/>)
- ⁵⁷ stanford.edu - ClinicalBertSum for text summarization, mentions ClinicalBERT, BioBERT. (<https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1204/reports/custom/report29.pdf>)
- ⁵⁸ arxiv.org - ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission. (<https://arxiv.org/abs/1904.05342>)

- ⁴ [pmc.ncbi.nlm.nih.gov](https://PMC11922739/) - Med-PaLM and Med-PaLM 2: architecture, training, evaluation, limitations. (<https://PMC11922739/>)
- ⁵ [researchgate.net](https://www.researchgate.net/publication/358938843_GatorTron_A_Large_Clinical_Language_Model_to_Unlock_Patient_Information_from_Unstructured_Electronic_Health_Records) - GatorTron: architecture, model sizes, training corpus, performance. (https://www.researchgate.net/publication/358938843_GatorTron_A_Large_Clinical_Language_Model_to_Unlock_Patient_Information_from_Unstructured_Electronic_Health_Records)
- ⁵⁹ [github.com](https://github.com/KentOn-Li/ChatDoctor) - ChatDoctor: base model (LLaMA), datasets, RAG with Wikipedia. (<https://github.com/KentOn-Li/ChatDoctor>)

引用的著作

1. 综述 : 人工智能在癌症诊断和治疗中的现状及潜在应用- 生物通, 访问时间为 六月 10, 2025, <https://www.ebiotrade.com/newsf/2025-5/20250521133401324.htm>
2. Deep Learning for Medical Image-Based Cancer Diagnosis - PMC - PubMed Central, 访问时间为 六月 10, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10377683/>
3. Large language models in healthcare - Paubox, 访问时间为 六月 10, 2025, <https://www.paubox.com/blog/large-language-models-in-healthcare>
4. Toward expert-level medical question answering with large ..., 访问时间为 六月 10, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11922739/>
5. (PDF) GatorTron: A Large Clinical Language Model to Unlock ..., 访问时间为 六月 10, 2025, https://www.researchgate.net/publication/358938843_GatorTron_A_Large_Clinical_Language_Model_to_Unlock_Patient_Information_from_Unstructured_Electronic_Health_Records
6. Anniversary Paper: History and status of CAD and quantitative ..., 访问时间为 六月 10, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC2673617/>
7. Holistic Evaluation of Large Language Models for Medical Applications | Stanford HAI, 访问时间为 六月 10, 2025, <https://hai.stanford.edu/news/holistic-evaluation-of-large-language-models-for-medical-applications>
8. Med-PaLM: A Medical Large Language Model - Google Research, 访问时间为 六月 10, 2025, <https://sites.research.google/med-palm/>
9. Revolutionizing Health Care: The Transformative Impact of Large Language Models in Medicine - PMC - PubMed Central, 访问时间为 六月 10, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11751657/>
10. The impacts of artificial intelligence in radiology - Marshall Digital Scholar, 访问时

间为六月 10, 2025,

<https://mds.marshall.edu/cgi/viewcontent.cgi?article=2948&context=etd>

11. AI in Healthcare Industry Adoption Challenges and Solutions - Amzur Technologies, 访问时间为六月 10, 2025, <https://amzur.com/blog/ai-in-healthcare-challenges-solutions>
12. Predicted Clinical Impact of Artificial Intelligence in Radiology: A Rapid Evidence Assessment - University of Malta, 访问时间为六月 10, 2025, https://www.um.edu.mt/library/oar/bitstream/123456789/132891/1/2519EMAEMA590705028413_1.PDF
13. Healthcare AI adoption up, but data and integration challenges persist, 访问时间为六月 10, 2025, <https://www.healthcareitnews.com/news/healthcare-ai-adoption-data-and-integration-challenges-persist>
14. Nuance DAX Infographic | Move Beyond Scribes, 访问时间为六月 10, 2025, <https://www.nuance.com/healthcare/dragon-ai-clinical-solutions/dax-copilot/infographic/move-beyond-scribes-to-automatically-document-care.html>
15. The Future of AI in Radiology with RamSoft, 访问时间为六月 10, 2025, <https://www.ramsoft.com/blog/future-of-ai-in-radiology>
16. Epic EHR AI: Generative AI Trends In 2025 To Transform Healthcare - SPsoft, 访问时间为六月 10, 2025, <https://spsoft.com/tech-insights/epic-ehr-ai-trends-in-2025-reshaping-care/>
17. Epic introduces Launchpad to fuel faster generative AI adoption among providers, 访问时间为六月 10, 2025, <https://www.fiercehealthcare.com/ai-and-machine-learning/epic-introduces-launchpad-fuel-faster-generative-ai-adoption-among>
18. How Shiv Rao's Abridge Is Freeing Clinicians from Administrative Burdens, 访问时间为六月 10, 2025, <https://www.massgeneralbrigham.org/en/about/newsroom/articles/how-shiv-raos-abridge-is-freeing-clinicians-from-administrative-burdens>
19. [2503.08420] Generalizable and Explainable Deep Learning for Medical Image Computing: An Overview - arXiv, 访问时间为六月 10, 2025, <https://arxiv.org/abs/2503.08420>
20. Revolutionizing Health Care: The Transformative Impact of Large Language Models in Medicine - Journal of Medical Internet Research, 访问时间为六月 10, 2025, <https://www.jmir.org/2025/1/e59069/>
21. A Review of Large Language Models in Medical Education, Clinical Decision Support, and Healthcare Administration - MDPI, 访问时间为六月 10, 2025,

<https://www.mdpi.com/2227-9032/13/6/603>

22. Nuance DAX Reviews, Pricing, Features & Integrations - Elion Health, 访问时间为六月 10, 2025, <https://elion.health/products/nuance-dax>
23. Advances in Image Processing and Pattern Recognition in Cancer Detection, Prediction, Diagnosis, and Prognosis - Preprints.org, 访问时间为六月 10, 2025, https://www.preprints.org/frontend/manuscript/11e4cc1244d3f2748f00570b992642a0/download_pub
24. Easing Workload Pressures While Maintaining Cancer Detection Accuracy - RSNA, 访问时间为六月 10, 2025, <https://www.rsna.org/news/2025/march/ai-as-a-second-reader-in-mammography>
25. AI In Medical Imaging Market Size & Outlook, 2030 - Grand View Research, 访问时间为六月 10, 2025, <https://www.grandviewresearch.com/horizon/outlook/ai-in-medical-imaging-market-size/global>
26. Artificial Intelligence in Radiology: Transforming Diagnostics and ..., 访问时间为六月 10, 2025, <https://www.ijsrtjournal.com/article/Artificial+Intelligence+in+Radiology+Transforming+Diagnostics+and+Raising+Ethical+Dilemmas>
27. U-Net Architecture For Image Segmentation | DigitalOcean, 访问时间为六月 10, 2025, <https://www.digitalocean.com/community/tutorials/unet-architecture-image-segmentation>
28. UNet++: A Nested U-Net Architecture for Medical Image Segmentation - PMC, 访问时间为六月 10, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC7329239/>
29. [2301.03505] Advances in Medical Image Analysis with Vision Transformers: A Comprehensive Review - arXiv, 访问时间为六月 10, 2025, <https://arxiv.org/abs/2301.03505>
30. [2211.10043] Vision Transformers in Medical Imaging: A Review - arXiv, 访问时间为六月 10, 2025, <https://arxiv.org/abs/2211.10043>
31. LLMs in Healthcare: Applications, Examples, & Benefits - AI21 Labs, 访问时间为六月 10, 2025, <https://www.ai21.com/knowledge/lms-in-healthcare/>
32. AI's Symbiotic Impact on Drug Development and Patient Experience in Global Pharma and Biotech | CustomerThink, 访问时间为六月 10, 2025, <https://customerthink.com/ais-symbiotic-impact-on-drug-development-and-patient-experience-in-global-pharma-and-biotech/>
33. Large Language Models in Healthcare: Uses, Benefits & More - Multimodal, 访问时间为六月 10, 2025, <https://www.multimodal.dev/post/large-language-models-in-healthcare>

34. Enhancing medical AI with retrieval-augmented generation: A mini narrative review - PMC, 访问时间为 六月 10, 2025,
<https://PMC12059965/>
35. Systematic Analysis of Retrieval-Augmented Generation-Based LLMs for Medical Chatbot Applications - MDPI, 访问时间为 六月 10, 2025,
<https://www.mdpi.com/2504-4990/6/4/116>
36. External Testing of a Commercial AI Algorithm for Breast Cancer Detection at Screening Mammography | Radiology: Artificial Intelligence - RSNA Journals, 访问时间为 六月 10, 2025, <https://pubs.rsna.org/doi/10.1148/ryai.240287>
37. Infervision Receives FDA 510(k) Clearance for Advanced Features in InferRead CT Lung, 访问时间为 六月 10, 2025,
[https://www.infervision.com.tr/en/news/infervision%2C-inferread-ct-akci%C4%9Ferindeki-geli%C5%9Fmi%C5%9F-%C3%B6zellikler-i%C5%87%C3%A7in-fda-510\(k\)-onay%C4%B1-ald%C4%B1](https://www.infervision.com.tr/en/news/infervision%2C-inferread-ct-akci%C4%9Ferindeki-geli%C5%9Fmi%C5%9F-%C3%B6zellikler-i%C5%87%C3%A7in-fda-510(k)-onay%C4%B1-ald%C4%B1)
38. U.S. FDA Grants Paige Breakthrough Device Designation for AI Application that Detects Cancer Across Different Anatomic Sites, 访问时间为 六月 10, 2025,
<https://www.paige.ai/press-releases/us-fda-grants-paige-breakthrough-device-designation-for-ai-application-that-detects-cancer-across-different-anatomic-sites>
39. The Paige Prostate Suite: Assistive Artificial Intelligence for Prostate Cancer Diagnosis - NCBI Bookshelf, 访问时间为 六月 10, 2025,
<https://www.ncbi.nlm.nih.gov/books/NBK608438/>
40. Abridge Alternative | AI Scribe for Clinical Conversations - RevMaxx, 访问时间为 六月 10, 2025, <https://www.revmaxx.co/abridge/>
41. Case Study: DAX Copilot - University of Michigan Health-West | Dictation Direct, 访问时间为 六月 10, 2025, <https://www.dictationdirect.com/case-studies/dax-copilot-university-of-michigan-health-west/>
42. Valley Hospital Reduces Burnout with DAX Copilot - Athenahealth, 访问时间为 六月 10, 2025, <https://www.athenahealth.com/resources/case-studies/valley-view-hospital>
43. Glass | AI Diagnosis & Clinical Decision Support (CDS), 访问时间为 六月 10, 2025, <https://glass.health/>
44. Glass.health - Callin.io, 访问时间为 六月 10, 2025, <https://callin.io/ai-tools/glass-health/>
45. Epic unveils new AI-enabled capabilities to improve EHR - Advisory Board, 访问时间为 六月 10, 2025, <https://www.advisory.com/daily-briefing/2025/03/14/epic-ehr-ec>

46. EHR vendors turn to artificial intelligence to modernize health records | Healthcare Dive, 访问时间为 六月 10, 2025,
<https://www.healthcaredive.com/news/ehr-artificial-intelligence-efforts-epic-oracle/741541/>
47. Hospitals are embracing AI-powered medical scribes - Mobius MD, 访问时间为 六月 10, 2025, <https://www.mobius.md/blog/hospitals-are-embracing-ai-powered-medical-scribes>
48. GatorTron: A Large Language Model for Clinical Natural Language Processing - medRxiv, 访问时间为 六月 10, 2025,
<https://www.medrxiv.org/content/10.1101/2022.02.27.22271257v2>
49. Generalizable and Explainable Deep Learning for Medical Image Computing: An Overview, 访问时间为 六月 10, 2025, <https://arxiv.org/html/2503.08420v1>
50. Large Language Models in Medicine: Clinical Applications, Technical Challenges, and Ethical Considerations - PMC, 访问时间为 六月 10, 2025,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC12086438/>
51. Impact of large language model (ChatGPT) in healthcare: an umbrella review and evidence synthesis - PMC, 访问时间为 六月 10, 2025,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC12057020/>
52. AI in Health Care and the FDA's Blinspot - Penn LDI, 访问时间为 六月 10, 2025,
<https://ldi.upenn.edu/our-work/research-updates/ai-in-health-care-and-the-fdas-blind-spot/>
53. Section 5 - Navigating the Regulatory Landscape: An Analysis of Legal and Ethical Oversight for Large Language Models (LLMs) | HIMSS, 访问时间为 六月 10, 2025, <https://www.himss.org/resources/section-5-navigating-regulatory-landscape-analysis-legal-and-ethical-oversight-large>
54. Timeline of large language models, 访问时间为 六月 10, 2025,
https://timelines.issarice.com/wiki/Timeline_of_large_language_models
55. (Open Access) BioBERT: a pre-trained biomedical language representation model for biomedical text mining. (2019) | Jinyuk Lee | 4625 Citations - SciSpace, 访问时间为 六月 10, 2025, <https://scispace.com/papers/biobert-a-pre-trained-biomedical-language-representation-4lpqv61ete>
56. Revolutionizing Biology Research With Lightning-Fast NLP: Introducing Sparse BioBERT, 访问时间为 六月 10, 2025,
<https://neuralmagic.com/blog/revolutionizing-biology-research-with-lightning-fast-nlp-introducing-sparse-biobert/>
57. ClinicalBertSum: RCT Summarization by Using Clinical BERT Embeddings - Stanford University, 访问时间为 六月 10, 2025,
<https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1204/reports/custom/r>

eport29.pdf

58. [1904.05342] ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission - arXiv, 访问时间为 六月 10, 2025, <https://arxiv.org/abs/1904.05342>
59. KentOn-Li/ChatDoctor - GitHub, 访问时间为 六月 10, 2025, <https://github.com/KentOn-Li/ChatDoctor>
60. The path forward for large language models in medicine is open - ResearchGate, 访问时间为 六月 10, 2025, https://www.researchgate.net/publication/386193568_The_path_forward_for_large_language_models_in_medicine_is_open
61. InferRead CT Lung - AI for Lung CT Scans - Infervision, 访问时间为 六月 10, 2025, <https://global.infervision.com/products/inferread-ct-lung>
62. Nanox scores clearance for cloud-connected X-ray bed after years of FDA review, 访问时间为 六月 10, 2025, <https://www.fiercebiotech.com/medtech/nanox-scores-clearance-cloud-connected-x-ray-bed-after-years-fda-review>
63. Zebra Medical Vision lands FDA clearance for tool to detect cardiovascular disease, 访问时间为 六月 10, 2025, <https://www.mobihealthnews.com/news/zebra-medical-vision-lands-fda-clearance-tool-detect-cardiovascular-disease>
64. United Imaging Healthcare Releases 2024 Annual and Q1 2025 Results, Highlights Strong Global Expansion and AI Innovation, 访问时间为 六月 10, 2025, <https://global.united-imaging.com/en/news-center/uih-news-list/2025-04-28>
65. Large Language Models Products - Elion Health, 访问时间为 六月 10, 2025, <https://elion.health/categories/large-language-models/products>
66. Abridge vs Scribe Health AI: Top AI Scribe & Note Taker for Your Business?, 访问时间为 六月 10, 2025, <https://www.scribehealth.ai/compare/abridge-alternatives>
67. 中国肺癌诊疗新范式：基于 AI 与分子可视化的早期筛查-精准诊断-个体化治疗全流程管理, 访问时间为 六月 10, 2025, <https://www.ebiotrade.com/newsf/2025-5/20250528074116309.htm>
68. AI in Medical Imaging Market Size Projected to Reach USD 14.46 Bn By 2034, 访问时间为 六月 10, 2025, <https://www.globenewswire.com/news-release/2025/02/13/3026027/0/en/AI-in-Medical-Imaging-Market-Size-Projected-to-Reach-USD-14-46-Bn-By-2034.html>
69. Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential, 访问时间为 六月 10, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC1955762/>

70. Trends, Challenges, and Applications of Large Language Models in Healthcare: A Bibliometric and Scoping Review - MDPI, 访问时间为 六月 10, 2025,
<https://www.mdpi.com/1999-5903/17/2/76>
71. Accuracy of Large Language Models When Answering Clinical Research Questions: Systematic Review and Network Meta-Analysis, 访问时间为 六月 10, 2025, <https://www.jmir.org/2025/1/e64486>